U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

PB-268 252

Classification of American Cities for Case Study Analysis. Volume II Detailed Report

Urban Systems Research and Engineering, Inc, Cambridge, Mass

Prepared for

Environmental Protection Agency, Washington, D C Office of Research and Development

Apr 77

PB 268 252

EPA-600/5-77-008b
April 1977

CLASSIFICATION OF AMERICAN CITIES FOR CASE STUDY ANALYSIS

Volume II

Detailed Report

By

Elizabeth Lake
Carol Blair
James Hudson
Richard Tabors
Urban Systems Research & Engineering Inc.
Cambridge, Massachusetts 02138

Contract #68-01-3299

Project Officer

Samuel Ratick
Office of Monitoring & Technical Support
Office of Research & Development
Washington, D.C. 20460

OFFICE OF MONITORING & TECHNICAL SUPPORT
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/5-77-008b	2.	3. RECIPIENT'S ACCESSION NO.
4. TITLE AND SUBTITLE Classification of American Cities For Case Study Analysis—Volume II Detailed Report		5. REPORT DATE April 1977 issuing date 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Elizabeth Lake, Carol Blair, James Hudson, Richard Tabors		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Urban Systems Research & Engineering Inc. 1218 Massachusetts Avenue Cambridge, Massachusetts 02138		10. PROGRAM ELEMENT NO. 1HA091 11. CONTRACT/GRANT NO. 68-01-3299
12. SPONSORING AGENCY NAME AND ADDRESS Office of Monitoring & Technical Support Office of Research and Development U.S. Environmental Protection Agency Washington, D.C. 20460		13. TYPE OF REPORT AND PERIOD COVERED Final 14. SPONSORING AGENCY CODE EPA/600/19

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Attempts to analyse and evaluate the impacts of federal programs has led to the extensive use of case studies of program impacts at selected sites. This project has developed a methodology for the systematic selection of representative case study sites and for generalising the study results. The methodology, involving two stage factor analysis and clustering, is applied to a specific program/policy problem, the selection of metropolitan areas for case studies in analysing the impact of federal policies on general environmental quality.

The methodology begins with a data base on standard metropolitan statistical areas SMSAs, including variables related to environmental quality, urban form, and household, industrial, and government activity. It analyses these variables through a two-stage factor analysis technique which allows heuristic consideration of the significant characteristics. Finally, it develops city clusters which group areas with similar attributes. Modal (or representative) cities are selected for each group and suggested as case study sites. These groups may be used to generalise the study results and to analyse the transferrability of results between areas. The methodology is sufficiently flexible to consider a wide range of research hypotheses.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS		b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Economic Factors	Socioeconomic Status	Factor Analysis	05F
Economic Surveys	Urban Areas	City Classification	
Census	Urban Sociology	Modal Cities	
Central City	Urban Geography		
Demographic Surveys			05C
Populations			05K

18. DISTRIBUTION STATEMENT Unlimited	19. SECURITY CLASS *(This Report)* Unlimited	21. NO.
	20. SECURITY CLASS *(This page)* Unlimited	22. PRICE A11-A01

EPA Form 2220-1 (9-73)

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

Attempts to analyze and evaluate the impacts of federal programs has led to the extensive use of case studies of program impacts at selected sites. This project has developed a methodology for the systematic selection of representative case study sites and for generalizing the study results. The methodology, involving two stage factor analysis and clustering, is applied to a specific program/policy problem, the selection of metropolitan areas for case studies in analyzing the impact of federal policies on general environmental quality.

The methodology begins with a data base on standard metropolitan statistical areas, SMSAs, including variables related to environmental quality, urban form, and household, industrial, and government activity. It analyzes these variables through a two-stage factor analysis technique which allows heuristic consideration of the significant characteristics. Finally, it develops city clusters which group areas with similar attributes. Modal (or representative) cities are selected for each group and suggested as case study sites. These groups may be used to generalize the study results and to analyze the transferrability of results between areas. The methodology is sufficiently flexible to consider a wide range of research hypotheses.

VOLUME II
CONTENTS

Abstract iii
Figures vii
Tables viii

1.0 Introduction 1

2.0 Research Design 3
  2.1 The Set of Localities to be Classified........ 3
  2.2 Relevant Attributes of an Urban Area........ 3
  2.3 Data Reduction and Clustering 10

3.0 Data Base 11
  3.1 Data Input to Stage I Factor Analysis........ 11
  3.2 Weaknesses of Data Collected........ 11

4.0 Stage I Factor Analysis........ 28
  4.1 SCB1: Environmental Quality........ 28
  4.2 SCB2: Urban Form and the Physical Environment 46
  4.3 SCB3: Pollutant Residuals 51
  4.4 SCB4: Household Sector 52
  4.5 SCB5: Public Sector........ 57
  4.6 SCB6: Industrial Sector 64

5.0 Stage II Factor Analysis 69
  5.1 Alternative Methods of Factor Rotation........ 69
  5.2 Selection of Stage I Factors to be Included in Stage II........ 72
  5.3 Results of Stage II Factor Analysis 73
  5.4 Analysis of Small, Medium and Large Cities 74

6.0 SMSA Groups Based on General Characteristics 80
  6.1 The Classification Process........ 80
  6.2 Input to the Grouping Algorithm 80
  6.3 SMSA Groups and Modal Cities 81
  6.4 Stability of the SMSA Groups 92

7.0 Relationship of the Basic SMSA Groups to Environmental Quality........ 105
  7.1 Regression Analysis 105
  7.2 Expanded Factor Analysis 112
  7.3 Significance Tests Between Groups 113

8.0 Classification of SMFAs for Alternative Purposes 122
  8.1 Potential Applications 122
  8.2 Transportation Analysis Application 123

Preceding page blank

CONTENTS (continued)

PAGE

Appendices:

A. Factor Analysis 139
B. Survey of City and Community Classifications 150
C. BMD Computer Package Description 180
D. Annotated Computer Output: Factor Analysis and Clusterings 183

FIGURES

PAGE

NUMBER

2-1 Boston SMSA 4
2-2 Causation: Environmental Quality 6
5-1 Alternative Methods of Factor Rotation 70

TABLES

Number		Page
3-1	Environmental Quality Variables (SCS1)	12
3-2	Urban Form and the Physical Environment Variables (SCS2)	13
3-3	Residual Pollutants (SCS3)	14
3-4	Household Descriptors (SCS4)	15
3-5	Public Sector Variables (SCS5)	16
3-6	Industry Variables (SCS6)	17
3-7	Effect of Missing Data on Sample Size (SCS1)	22
3-8	SMSAs for Which Data is Systematically Missing	24
4-1	SCS1 Missing Variables	30
4-2	Environmental Quality: Missing Data	31
4-3	Distribution of Manganese Values (Observed	43
4-4	Stage I Factor Analysis - SCS1	44
4-5	Distribution of Environmental Quality Variables	45
4-6	Indicator Variables	46
4-7	SCS2 - Stage I - Initial Factor Analysis	48
4-8	SCS2 - Stage I - Factor Analysis of 15 Variables	50
4-9	SCS3 - Stage I - Final Factors	53
4-10	SCS4 - Stage I - Initial Factors	54
4-11	SCS4 - Stage I - Second Iteration Factors	56
4-12	SCS4 - Stage I - Test of the Effect of City Size	58
4-13	SCS4 - Stage I - Factors	59
4-14	SCS4 - Stage I - Interpretation of Final Factors	60
4-15	SCS5 - Stage I - First Iteration	62
4-16	SCS5 - Stage I - Final Factors	63
4-17	SCS6 - Stage I - Initial Factor Analysis	65
4-18	SCS6 - Comparison of SEAS and Census Industry Data	66
4-19	SCS6 - Stage I - Final Factors	68
5-1	Preliminary Stage II Factors	71
5-2	Proportion of Variance Explained by Stage I Factors	74
5-3	Input to Stage II Factor Analysis (Stage I Factors)	75
5-4	Stage II Factors	76
6-1	Groups of Similar SMSAs	82
6-2	Comparisons Between Selected Groups and Subgroups	84
6-3	Characterization of SMSA Groups Through Stage II Factors	85
5-4	Characterization of SMSA Groups by Stage II Factors	86
6-5	Treatment of Missing Data	88
6-6	Centroids of Major Groups from Four and Six Factors	93
6-7	Four Factor Classification and Six Factor Classification	94
6-8	Classification of Small SMSAs	96
6-9	Classification of Medium SMSAs	99
6-10	Classification of Large SMSAs	102
6-11	Classification of Small, Medium and Large SMSAs	104

TABLES (continued)

Number Page

7-1 Integration of Environmental Descriptors with SMSA Classification .. 105
7-2 Regression Statistics 106
7-3 Regression Analysis; DO Against Fifteen Stage I Factors 108
7-4 Regression Analysis: BOD Against Fifteen Stage I Factors 108
7-5 Regression Analysis: TURB Against Fifteen Stage I Factors 109
7-6 Regression Analysis: HARD Against Fifteen Stage I Factors 109
7-7 Regression Analysis: PHS Against Fifteen Stage I Factors 110
7-8 Regression Analysis: SO2 Against Fifteen Stage I Factors 110
7-9 Regression Analysis: SUS Against Fifteen Stage I Factors 111
7-10 Regression Analysis: PDI Against Fifteen Stage I Factors 111
7-11 Mean Values of Environmental Indicators by Group 115
7-12 Comparison of SMSA Groups Using DO as the Measure 116
7-13 Comparison of SMSA Groups Using BOD as the Measure 116
7-14 Comparison of SMSA Groups Using HARD as the Measure 117
7-15 Comparison of SMSA Groups Using PHS as the Measure 117
7-16 Comparison of SMSA Groups Using TURB as the Measure 118
7-17 Comparison of SMSA Groups Using SUS as the Measure 118
7-18 Comparison of SMSA Groups Using SO2 as the Measure 119
7-19 Comparison of SMSA Groups Using PDI as the Measure 119
7-20 Levels of Environmental Quality by SMSA Group 121
7-21 Comparison of City Groups Over all Environmental Variables 121
8-1 Variables Selected for the Transportation Classification 123
8-2 Means for Variables by City Size Group 125
8-3 Factor Analysis for Large Cities 126
8-4 Factor Analysis for Medium Cities 127
8-5 Factor Analysis for Small Cities 128
8-6 Factor Analysis for All Cities 129
8-7 Factors Relevant to Urban Transportation Analysis 130
8-8 Classification of Large Cities 132
8-9 Classification of Medium Cities 133
8-10 Classification of Small Cities 133
8-11 Classification of Large Cities 134
8-12 Classification of Medium Cities 135
8-13 Classification of Small Cities 136

CLASSIFICATION OF AMERICAN CITIES FOR CASE STUDY ANALYSIS

VOLUME II

DETAILED REPORT

1.0 INTRODUCTION

The research described in this report has developed two major products, one a direct output, and the other a methodology for further analysis. The first output has been a classification of Standard Metropolitan Statistical Areas (SMSAs) based on broad measures of environmental quality and other attributes. This classification depends on a large scale data base which includes 262 SMSAs, and has data on activity in the industrial, demographic, and government sectors, on the attributes of the urban form and the physical environment, and on the pollutant residuals and ambient environmental quality resulting from those activities and attributes. The second product is a methodology for developing alternative classifications, oriented toward specific policy or research issues, and the urban characteristics related to them. The methodology can be directed at issues such as the choice of sites for case studies and demonstration projects, or the transfer of results from one case study area to other cities.

To develop the general SMSA classification scheme, a two stage factor analysis technique was applied to the data base. In Stage I, a small number of factors was extracted from the data in each SCS. These factors summarized the basic dimensions of the data available to describe relevant attributes of urban areas. In the second stage, factors derived in Stage I were treated as variables. The factors derived in Stage II, then, reflected relationships both within the SCSs and between SCSs. The four factors from Stage II which explained the greatest amount of variance in the data base were taken to characterize the SMSAs for purposes of classification. As the factors generated are linear combinations of the original variables, it is possible to estimate scores for each observation on each factor. These provide the location of each SMSA in the four-dimensional factor space derived in Stage II. (The 263 SMSAs have been classified by applying a simple "nearest neighbor" clustering technique to these factor scores.) Five major city groups were developed, they included 175 of the 263 SMSAs.

Other classifications might be developed to satisfy a more specific emphasis. This research included such an effort for the Energy Research Development Agency, interested in potential energy savings through changes in transportation patterns. The basis of the classification was limited to variables related to auto use: auto ownership, per capita vehicle-miles travelled, household size, urban density, etcetera. Comparison of the resulting groups with those previously developed indicates a great deal is common to the two classifications. The data bank, the methodology, the classification and modal cities will be valuable in a variety of applications related to urban development and environmental quality.

The research design is discussed in Chapter 2; the variables contained in the data base together with their strengths and limitations are described in Chapter 3. First and second stage factor analyses are summarised in Chapters 4 and 5 respectively; while the city groupings obtained on the basis of the second stage factors are described in Chapter 6. The applicability of these groups and their modal cities to general environmental research are analysed in Chapter 7. Specific applications of the site selection capability are discussed in Chapter 8.

2.0 RESEARCH DESIGN

The fundamental objective of this research was to group together metropolitan areas which are similar in terms of both environmental quality and factors which affect environmental quality. Essential to the design of any classification methodology are (1) definition of the entities to be classified, (2) identification of attributes to be considered in the classification, and (3) selection of appropriate techniques by which the data can be used together to differentiate between observations.

2.1 The Set of Localities to be Classified

For this research, the definition of entities to be classified was a simple matter. Standard Metropolitan Statistical Areas (SMSAs) have been defined by the Office of Management and Budget (OMB) to represent the areas in and around one or more cities, which act as the center, in which the activities form an integrated economic and social system. Although other definitions of U.S. metropolitan areas are available, none have been utilized as extensively as the SMSA for the collection of data. The use of SMSAs achieved the greatest possible amount of data consistent between metropolitan areas of the U.S.

In this report, the two terms--cities and SMSAs--are used interchangeably. Cities, however, make up parts of SMSAs, as demonstrated in Figure 2-1.

Cities are constantly changing and to reflect this change OMB revises the definition of an SMSA periodically, therefore, it was necessary to perform the classification from the perspective of a single point in time. The set of SMSAs defined as of January 7, 1972 was selected as the set of cities to be classified and data utilized in the classification was that which most closely represented each city at that point in time.

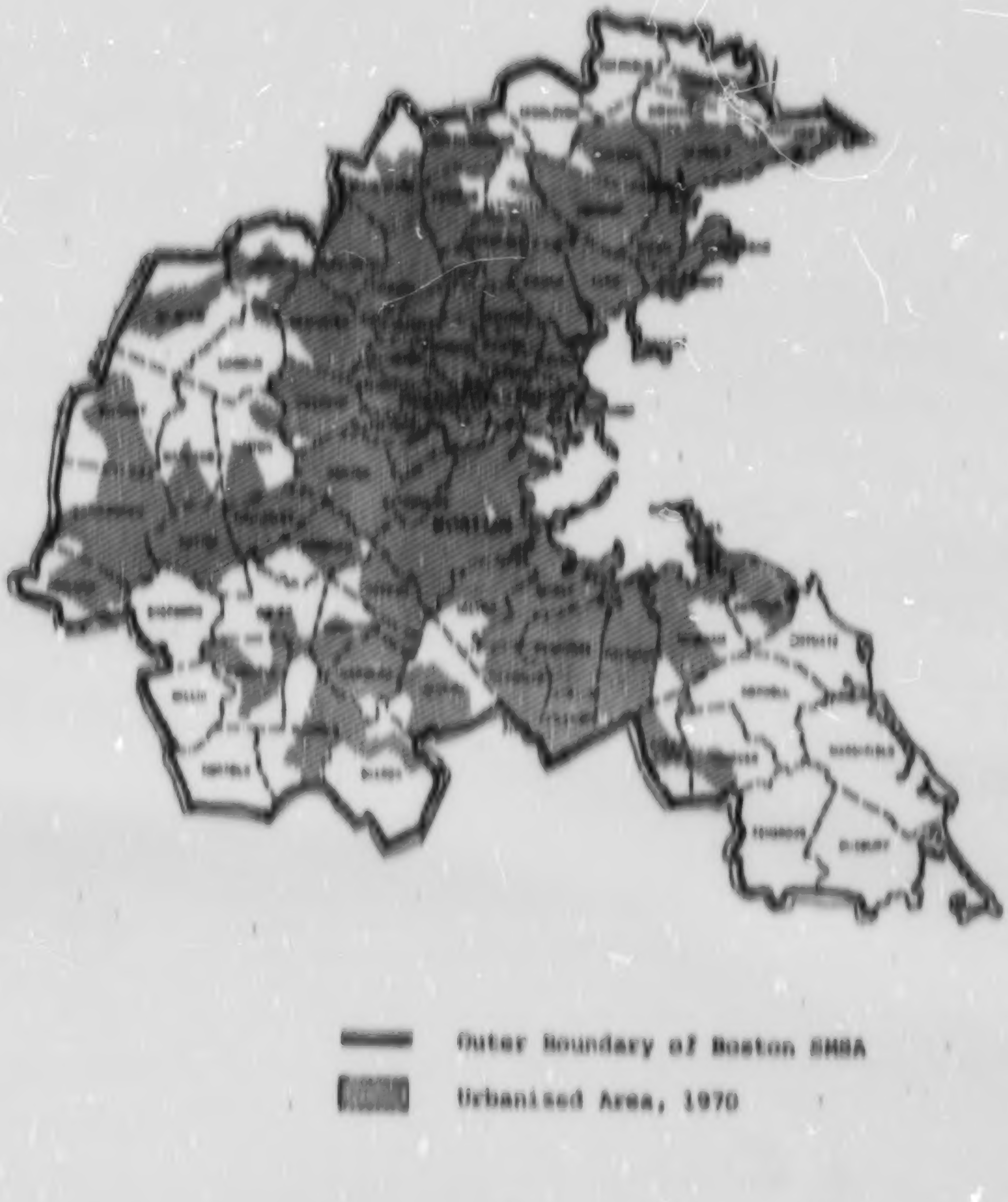
2.2 Relevant Attributes of an Urban Area

The set of measures which reflect environmental quality and its determinants is large and diverse since almost all systems of an urban area either affect or are affected by environmental quality. The major determinants and effects are demonstrated in the simplistic diagram of Figure 2-2.

Beginning at the bottom of the diagram, ambient environmental quality is shown to be a function of both residuals and the capacity of the environment to dilute and/or neutralize pollutants (urban form and physical environment). Urban form also influences residuals generation, particularly in the transportation sector. Residuals are the net result of pollution generated by households and industry and abatement efforts by the public sector. The public sector also influences consumption through investment in public facilities. However, the three areas at the top are reduced to simple form by attributing responsibility for abatement with the public sector, consumption with households, and production with industry. Second order effects, such as the influence of environmental quality on the three actors at the top of the causal path, are not represented here. The primary areas of interest, then, are:

Figure 2-1

Boston SMSA



— Outer Boundary of Boston SMSA

▨ Urbanised Area, 1970

1. ambient environmental quality
2. urban form and the physical environment
3. residuals
4. demographic characteristics
5. government
6. industry.[*]

Each of these areas has been designated a significant characteristic set (SCS) to signify a group of attributes describing a particular aspect of the urban area. The SCSs provide a vehicle through which research hypotheses have been articulated and a convenient framework for the collection, documentation and analysis of relevant data.

Expression of the research hypotheses is particularly important to this research since this approach to city classification is different from any previous work. Each SCS is discussed separately below.

Every effort has been made, in selecting appropriate measures for the attributes, to eliminate the effect of city size which tends to dominate any classification. In most cases, per capita measures were utilized.

2.2.1 Environmental Quality --

Metropolitan area environmental quality is a complex, nonlinear result of the activities in the area, their spatial location, and the assimilative capacity of the natural environment. Urban form and governmental policies both have a major impact on environmental quality.

However, the environment causes changes in the metropolitan region, through its effects on housing values,[**] recreation,[***] and public health.[†] The amount and potential for excess capacity in the drinking water system is

---

[*] Although environmental quality is closely related to energy production and consumption, this framework does not include a special category for energy variables. Energy production takes place in the industry sector and energy consumption is by the industry and household sectors. Therefore, a complete description of industry and household activities should explain any variance in environmental quality between SMSAs which is related to energy measures.

[**] D.M. Dornbusch & S.M. Barrager, Benefit of Water Pollution Control on Property Values, prepared for U.S. Environmental Protection Agency (Washington, D.C., 1973).

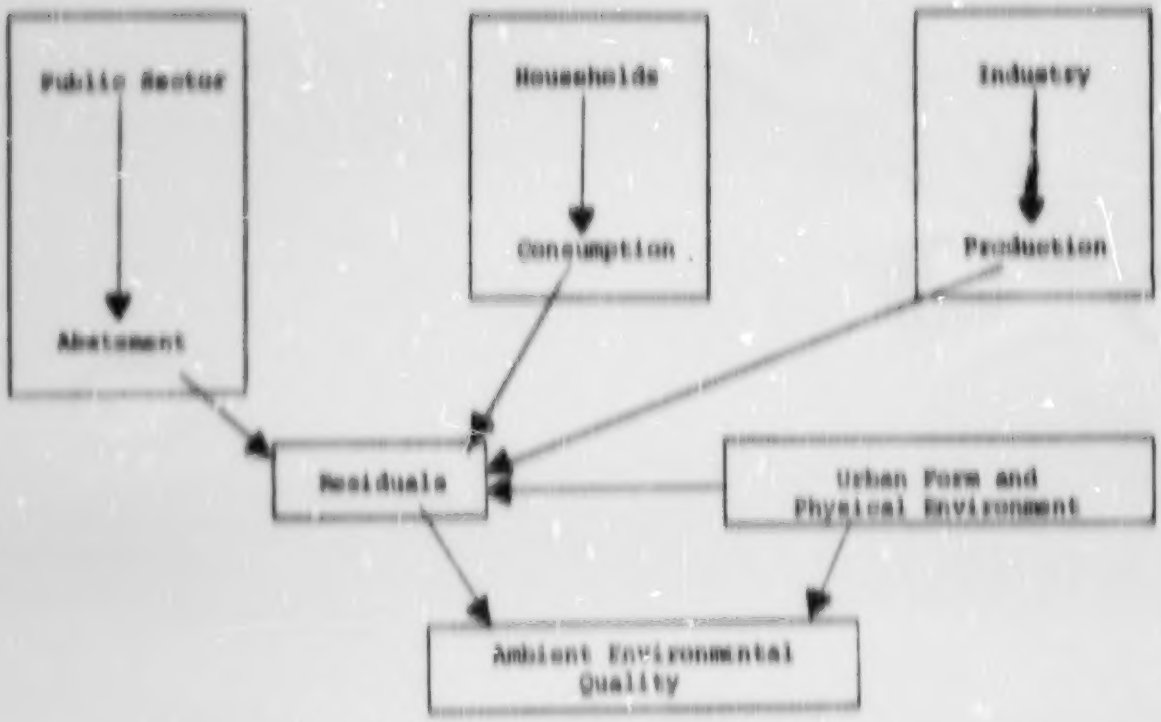
[***] Urban Systems Research & Engineering, Inc., The Recreation Benefits of Water Quality Improvement: Analysis of Day Trips in an Urban Setting, prepared for the U.S. Environmental Protection Agency (Washington, D.C., 1975).

[†] U.S. Water Resources Council, The Nation's Water Resources (Washington, D.C., 1968).

Figure 2-2

Causation: Environmental Quality

(First Order Effects)


Public Sector
Abatement
Households
Consumption
Industry
Production
Residuals
Urban Form and
Physical Environment
Ambient Environmental
Quality

also an important determinant of urban growth. To measure the interactions between overall ambient environmental quality and other aspects of the urban area would be impossible at this point. However, the classification of metropolitan areas by all these attributes together is a promising approach.

Environmental quality is represented in this SOS by ambient water quality, ambient air quality, drinking water quality, and available subjective measures. Drinking water quality is related to environmental quality although, in most cases, the water undergoes extensive treatment before distribution for consumption. Subjective measures, although not widely available or precise, add another dimension to the prospective. Noise, although a significant aspect of the environment, was recognised from the beginning as an attribute on which insufficient uniform data would be available. Measures of ambient air and water quality, then, are assumed to be the primary indicators of environmental quality.

2.2.2 Urban Morphology and Location --

The spatial structure and location of an SMSA can play a major role in determining its environmental quality. Compact, densely populated areas are more easily served by sewerage lines but frequently require larger treatment facilities with greater point source effluent discharges. In much the same way, dense urban areas are subject to vehicle congestion causing increased emissions per unit area though distances covered per vehicle are likely to be less, again somewhat offsetting environmental impacts.

The relationship between urban form and environmental quality has been recognised for a considerable period of time though the problems of attaching cause and effect have been considerable. Berry, et al. summarise their findings in light of the previous experience:

> "What we have added to the previous literature is firm evidence that urban form plays a significant role in translating these basic city characteristics (size, density, and the economic base) into land use, and that the land use pattern, as shaped by urban form, is directly related to the nature and intensity of environmental pollution. Most importantly, holding the effects of city size and manufacturing concentration constant, and controlling for an inverse relationship between income levels and pollution (probably an effect of pollution rather than a cause, via migration of people and activities to the least polluted amenity-rich environments in the past two decades), we have found that:
>
> 1) The core-oriented urban region with a radial transportation network and a steep density gradient
>   (a) displays greater intensity of land use, a lower percentage of land developed and used for residential and commercial purposes, and more open space, and
>   (b) as a consequence of this land use mix and pattern, has superior air and water quality to:
>
> 2) The dispersed urban region, which has a less focused transport network and lower, more uniform population densities. This urban form

(a) displays urban sprawl, with a higher percentage of residential and commercial land use and less open space than in the core-oriented case, and

(b) as a consequence of this land use mix, has inferior air and water quality.*

Our general hypothesis concerning the relationship between environmental quality and urban morphology is that the spatial distribution of population and economic activity determined, in part, the observed difference in environmental quality between metropolitan regions. We expect to be able to separate the more dense, generally radial urban centers of the Northeast from the more dispersed areas of the South and West. We expect to be able to separate the transport "hub" cities from more traditional manufacturing and service centers and we expect to be able to separate those dense urban areas with highly developed wastewater systems from those less dense centers with less development of centralized wastewater treatment facilities. The complex interactions between these variables are a function of not only the pollutant load of any population, or its economic activity, but also the location and extent of these activities. It is not enough to know the area, in square miles, of an SMSA, it is also necessary to know the relative shape and distribution of the population within the SMSA, the amount of area used for industrial purposes, for residential use, and for open space.

In addition to characterizing urban form, this SCS reflects the assimilative capacity of the physical environment, which will clearly affect environmental quality. In terms of water quality, for instance, higher temperatures mean increased organic activity and cleaning; higher rainfall means greater diffusion of pollution and also increased runoff; and hardness and dissolved oxygen affect the water's ability to assimilate a variety of pollutants. Weather factors affect air quality in a similar manner.

These aspects of urban form and assimilative capacity together include the major causes of variation in environmental quality.

2.2.3 Residual Pollutants --

Residuals measure the pollutant load placed on the environment by activities in the area. Residuals are defined as nonproduct outputs, the spent hops in the brewing process are such residuals. Gross residuals measure the total amounts of such outputs. Some proportion of this is captured through abatement activities, the remainder is defined as the net residuals. The magnitude and composition of these pollutants should be consistent with activities measured by other SCSs, with environmental quality measured by other SCSs, and with environmental quality as measured by SCS1.

2.2.4 Households --

The amount of consumption, and its composition, will affect the generation of residuals and, thus, environmental quality. Food consumption, for example,

---

*Brian J.L. Berry, et al. *Land Use, Urban Form and Environmental Quality*, University of Chicago, Department of Geography Research Paper No. 155, pp. 411-412.

results in more wastes than personal consumption expenditures on services--such as theaters or concerts. The amounts of consumption and residual generation are positively correlated, while other things being equal, the quantity of consumption has an inverse relationship with ambient environmental quality.

Some differences in the consumption patterns of SMSAs will be explained by demographic variables. Population density, family size, the age, and other attributes of the population will clearly affect both the volume and composition of consumption. Growth and income measures are also useful in explaining the differences in consumption patterns between SMSAs. Income is relevant because the propensity to consume varies inversely with income levels, and because the composition of consumption is a function of the level of income. The composition of housing will also influence the environmental quality of SMSAs. It has been argued that single family dwellings generate more residuals and runoff than the more concentrated residential forms. All these descriptors of the population and its activities are related to the generation of pollutants within the SMSA.

While environmental quality is affected by consumption levels and patterns, it also affects the structure of personal consumption expenditures. High levels of air pollution necessitate extra cleaning, painting, medical expenses, and quality deterioration implies higher costs for the production of potable water.

2.2.5 Public Sector --

Governmental activities in metropolitan areas have both straightforward and complex relationships with environmental quality. For example, expenditures on abatement facilities tend to reduce the residuals load entering the environment. Regulations such as requirements for emission devices on automobiles, reduce air pollution. User charges and pretreatment requirements for industries using publicly-owned wastewater treatment facilities tend to modify the behavior of the private sector in such a way that residuals generation is reduced. Governmental activities are influenced by the desires of the electorate. If citizen and conservation groups voice dissatisfaction with the environment, the local government may be persuaded to step up its abatement and regulatory activities.

2.2.6 Industry --

As production generates residuals, in the simplest case, it may be hypothesized that more production will lead to the generation of more residuals. The structure of productive activity is important as well, since the pollution intensity of the various productive sectors differs greatly. Of nine sectors often utilized, mining, manufacturing, transportation, and public utilities generate far more pollution than finance, insurance and real estate, or the government. In other words, the pollutant generating coefficients vary between sectors. If an SMSA's productive activities are dominated by sectors which produce few residuals, its ambient air and water quality may be expected to be relatively better than if the industrial mix contained the more intensive residual generates. Ambient quality, then, depends on the sectoral composition of productive activity.

2.3 Data Reduction and Clustering

Sequential application of the factor analysis technique allows the reduction of large numbers of data variables into a small set for grouping observations. The factor analysis technique was chosen because it can simplify a data base by reducing the number of parameter dimensions involved. This transformation is made possible when there is moderate or high correlation between sets of variables which compose the original data base; when two or more variables are very highly correlated, a single "factor" may describe most of their variation or correlation adequately. In geometric terms, correlation between original variates indicates nonorthogonal dimensions which can be simplified in the vector space.* The input to the factor analysis was designed according to the research hypotheses to insure that the results provide meaningful factors which identify the basic dimensions of the data.

In the most standard use of factor analysis, the variables chosen are assumed to be the universal set; i.e., the variables include all information relevant to the problem under study. Generally, no attempt is made to analyze categories within the data set. For this reason, most variables are selected on general relevancy and availability criteria. The result of this is an urban typology, the specific characteristics of which cannot be disaggregated or interpreted. It is not verifiable in the sense that only with difficulty does the analyst know what individual sector patterns went into the creation of the grouping. Even if a sectoral analysis is done ex post facto, the ability to change the relative importance or weight of the axes is greatly limited.

The two-stage application of the factor analytic technique was utilized here to insure the proper evaluation of research hypotheses. Stage I analysis involved a separate factor analysis for each SCS; to reveal the hypothesized underlying dimensions within each group of variables. The output from all Stage I analyses was then combined and used as the input to Stage II. Stage II thus identified the relationship between SCSs and the basic underlying attributes of U.S. metropolitan areas.

The identification of modal cities was accomplished by describing groups of similar cities and then choosing the most typical city from each group. The important part of this process is clustering the observations. Factor scores from Stage II above describe the 262 SMSAs in a multidimensional space; therefore, the grouping can be based entirely on the relative positions of the observations (SMSAs) in that space. Cities were classified in just this manner; an algorithm is used which considers each city a group initially. The two groups separated by the smallest geometric distance are then located and grouped together, then the two nearest of the new set of groups are combined, and so the process continues until only one group remains. The grouping process is then retraced to determine the stage at which the optimal level of grouping has been achieved. Criteria for choosing a stopping point include the size and number of groups and the relationship of variance explained by the clusters to total variance in the data at any point. Once the preferred set of clusters is chosen, the modal city is identified by simply selecting the city in each group which lies closest in the multidimensional space to the geometric centroid of the group.

---

*See Appendix A for a comprehensive discussion of the factor analysis technique.

3.0 DATA BASE

Volume III of this report presents a complete documentation of data used in this research, including a listing. A summary description of the data set and some of the problems with the data is presented here for convenience. A huge amount of data was located in the collection process. Of this body of data, only a certain amount was found to be sufficiently uniform in meaning and format and to include a large enough number of SMSAs to be potentially useful to this research. A list of the measures actually input to Stage I factor analysis is given below, followed by a discussion of the weaknesses of this subset of data.

3.1 Data Input to Stage I Factor Analysis

With the shortcomings indicated below and minor problems outlined in Volume III, the data collected is still the best available at the time. The quality of the data should be considered at every point in the following analysis. Six tables (Tables 3.1 - 3.6) are presented here to identify the data base which was utilized in the analysis. Each table pertains to one SCS.

3.2 Weaknesses of the Data Collected

In processing the data and performing preliminary factor analysis, a great deal was learned about the data which had been collected. The following discussion identifies the inherent inadequacies of the data set with respect to the research hypotheses and problems with data collected.

3.2.1 Relationship of the Data to Research Hypotheses --
The research hypotheses set forth in Section 2.2 of this volume cannot be completely represented due to a lack of usable data measuring a number of attributes of an urban area. The data for five of the six SCSs show weaknesses in this respect.

SCS1 is intended to describe overall environmental quality--yet a sufficient number of observations was available for only two measures of air quality--suspended particulates, and sulfur dioxide. National Ambient Air Quality Standards have been established for five "criteria" pollutants: suspended particulates, sulfur dioxide, carbon monoxide, oxidants, and nitrogen dioxide. The proportion of SMSAs with missing data is shown below for each measure.

Suspended particulates	2%
Sulfur dioxide	12%
Oxidants	55%
Carbon monoxide	62%
Nitrogen dioxide	very little data available.

The first two measures, having relatively complete data, were utilized in the classification to represent overall environmental quality.

Table 3-1

Environmental Quality Variables (SCS1)

	Code	Number of Observations
A. Ambient Water Quality		
Biochemical Oxygen Demand (5-day, 20°C)	BOD	231
Fecal Coliforms, measured by membrane filter method	FCOL	213
Total Nitrogen (mg/l)	N	172
Total Organic Carbon (mg/l)	C	205
Dissolved Solids (mg/l)	TDSS	225
Suspended Solids (mg/l)	TSS	220
Turbidity (Jackson Candle Units)	TURB	228
Alkalinity (mg/l as $CaCO_3$)	AALK	251
Acidity (standard units)	PHS	255
Oil and Grease, soxhlet extraction (mg/l)	OAG	149
Methylene Blue Active Substance (mg/l)	MBAS	161
B. Ambient Air Quality		
Suspended Particulates (micro-g/cu.m.)	SUS	256
Sulfur Dioxide (micro-g/cu.m.)	SO2	233
C. Subjective Measures		
PDI Index	PDI	342
D. Drinking Water Quality		
Chloride (mg/l)	CL	207
Fluoride (mg/l)	FL	202
Iron (mg/l)	FE	206
Manganese (mg/l)	MG	201
Nitrate (mg/l)	NO3	196
Sulfate (mg/l)	SO3	192
Alkalinity (mg/l as $CaCO_3$)	ALK	189
Hardness (mg/l as $CaCO_3$)	HD	191
Acidity (pH standard units)	PH	204
Total Dissolved Solids (mg/l)	TDS	188

Table 3-2

Urban Form and the Physical Environment Variables (SCS2)

	Code	Number of Observations
A. Size		
Square Miles per Person -- SMSA	SQMS	262
Square Miles Per Person -- Central City	SQMC	262
Square Miles Per Person -- Urban Places	SQMU	235
B. Distribution of Activity		
Portion of Total Population in Central City	CENP	262
Portion of Manufacturing in Central City (by value added) (percent)	CENM	231
Portion of Retail Sales in Central City (percent)	CENS	256
Portion of Employment in Central City (percent)	CENE	262
Portion of Manufacturing Employment in the Central City (percent)	CENME	262
C. Number of Population Centers		
Arc of SMSA around the Center (quadrants)	ARC	262
Number of Major Radial Roadways	RAD	262
Number of Circumferential Roadways	CIR	262
Portion of Workers Working Outside SMSA (percent)	TCOM	248
Total Miles of Roadway per Capita	HMPC	228
Portion of Principal Arterials (percent)	HMPR	228
D. Open Space		
Portion of Land Which is Urbanized (percent)	LU	236
Portion of Land Devoted to Outdoor Recreation (percent of total land area)	REC	239
Land in Farms (percent)	LA	240
E. Assimilative Characteristics		
Latitude of SMSA (degrees)	LAT	262
Longitude of SMSA (degrees)	ALT	262
Altitude (Feet)	ALT	262
Mean Annual Precipitation (inches)	PRE	262
Mean Annual Possible Sunshine (percent)	SUN	136

(continued on following page)

Table 3-2 (continued)

	Code	Number of Observations
Mean Annual Wind Velocity (miles per hour)	WIND	191
Inversions (mean annual frequency)	INV	242
Water Temperature (ambient) (°C)	WTMP	257
Dissolved Oxygen in Water (ambient)(mg/l)	DO	246
Hardness of Water (ambient) (mg/l as $CaCO_3$)	HARD	249
Large Water Bodies (number)	WTR	262

Table 3-3

Residual Pollutants (SCB3)

	Code	Number of Observations
A. Air		
Particulates (tons per year per capita)	RPAR	249
Sulfur Oxides (tons per year per capita)	RSO	249
Nitrogen Oxides (tons per year per capita)	RNOX	249
Hydrocarbons (tons per year per capita)	RHC	249
Carbon Monoxide (tons per year per capita)	RCO	249
B. Water		
Biochemical Oxygen Demand (tons per year per capita)	RBOD	249
Suspended Solids (tons per year per capita)	RSS	249
Dissolved Solids (tons per year per capita)	RDS	249
Nutrients (tons per year per capita)	RNUT	249
Wastewater (million gallons per year per capita)	RWW	249
C. Land		
Noncombustible Solid Waste (tons per year per capita)	RNSW	249
Industrial Sludges (tons per year per capita)	RIS	249

Table 3-4

Household Descriptors (SC84)

	Code	Number of Observations
A. Social Status		
Median Age (years)	AGE	242
Portion of Population which is of Foreign Stock (percent)	FF	262
Fertility Rate (children ever born per thousand women ever married)	RF	262
Married Couples without own Household (percent)	MOH	242
Portion of Population which is Urban (percent)	PU	262
Individuals Residing in the Same Dwelling for Five Years (percent)	NMOV	262
Average Household Size (persons per household)	PPH	260
Portion of Population which is Black (percent)	PBL	257
Movers into the SMSA (percent of individuals over five years of age)	MIGF	262
Female/Male in 20-64 Age Group	FTM	262
B. Housing		
Single Family Dwellings (percent)	H1U	258
Owner Occupied Housing (percent)	HOCC	258
Units in Structure--More than Five Units (percent)	H5U	236
Units in Structure--More than Fifty Units (percent)	H50U	235
C. Growth		
Population Change -- Central City (percent)	PCC	260
Population Change -- SMSA (percent)	PCS	258
Change in Median Income (percent)	YMC	222
D. Economic Welfare		
Income -- Gini Coefficient	GINI	262
Median Family Income (dollars)	YM	262
Per Capita Income (dollars)	YP	262
Employment/Population -- Male (percent)	EPM	262

(continued on following page)

Table 3-4 (continued)

	Code	Number of Observations
Employment/Population -- Female (percent)	EPF	262
Employment/Population -- Total (percent)	EPT	262
Employment/Population -- Black (percent)	EPB	257
E. Living Conditions		
Crowded Housing Units (percent)	HC	262
Relative Death Rate (adjusted for age) (deaths per thousand population)	DR	262
Infant Death Rate (deaths per thousand)	DI	242
F. Modal Split		
Households with One or More Automobiles (percent of total households)	AUTO	262
Work Trip -- Drivers (percent)	TD	242
Work Trip -- Passengers (percent)	TP	242
Work Trip -- Walkers (percent)	TWK	242
Vehicle-Miles Travelled (per capita)	VMTP	224
G. Degree Days		
Mean Temperature in January (°F)	TJA	262
Mean Temperature in June (°F)	TJU	262

Table 3-5

Public Sector Variables (SC85)

	Code	Number of Observations
A. Expenditures Related to the Environment		
Sewerage Expenditures -- Total (dollars per capita)	GTS	262
Sewerage Expenditures -- Other than Capital Outlay (dollars per capita)	GXSM	262
Expenditures on Sanitation other than Sewerage (dollars per capita)	GXSO	261
Expenditures on Water Supply (dollars per capita)	GXW	259

(continued on following page)

Table 3-5 (continued)

	Code	Number of Observations
Expenditures on Parks and Recreation (dollars per capita)	GXPR	262
B. Community Concern And Involvement		
Presidential Votes Cast Per Voting Age Population (votes per person)	VOT	242
Sunday Newspaper Circulation per Capita (papers published per thousand population)	NEW	241
Local Government -- General Expenditures (dollars per capita)	GX	259
Local Government -- General Revenue (dollars per capita)	GREV	259
Local Government -- Employment (full time equivalent per capita)	GEMP	259

Table 3-6

Industry Variables (SC86)

	Code	Number of Observations
A. Industries Critical to Environmental Quality		
Total Value of Production in Meat Animals and other Livestock (dollars per capita)	IFLV	248
Total Value of Production in Crude Petroleum, Natural Gas (dollars per capita)	IOIL	248
Total Value of Production in Meat Products (dollars per capita)	IMET	248
Total Value of Production in Broad and Narrow Fabrics (dollars per capita)	ICTN	248
Total Value of Production in Household Furniture (dollars per capita)	IFRN	248
Total Value of Production in Pulp Mills (dollars per capita)	IPLP	248
Total Value of Production in Paper and Paperboard Mills (dollars per capita)	IPPR	248

(continued on following page)

Table 3-6 (continued)

	Code	Number Observations
Total Value of Production in Industrial Chemicals (dollars per capita)	ICHM	248
Total Value of Production in Commercial Printing (dollars per capita)	IPRT	248
Total Value of Production in Fertilizers (dollars per capita)	IFRT	248
Total Value of Production in Miscellaneous Chemical Products (dollars per capita)	IMCM	248
Total Value of Production in Plastic Materials and Resin (dollars per capita)	IPLA	248
Total Value of Production in Paints (dollars per capita)	IPNT	248
Total Value of Production in Petroleum Refining (dollars per capita)	IFUL	248
Total Value of Production in Paving and Asphalt (dollars per capita)	IASP	248
Total Value of Production in Glass (dollars per capita)	IGLS	248
Total Value of Production in Structural Clay Products (dollars per capita)	ICLY	248
Total Value of Production in Cement, Concrete, Gypsum (dollars per capita)	ICMT	248
Total Value of Production in Steel (dollars per capita)	ISTL	248
Total Value of Production in Aluminum (dollars per capita)	IALM	248
Total Value of Production in Household Appliances (dollars per capita)	IAPL	248
Total Value of Production in Motor Vehicles (dollars per capita)	ICAR	248
Total Value of Production in Electric Utilities (dollars per capita)	IELC	248
Total Value of Production in Coal Mining (dollars per capita)	ICOL	248
Total Value of Production in Canned and Frozen Foods (dollars per capita)	IVEG	248

(continued on following page)

Table 3-6 (continued)

	Code	Number of Observations
Total Value of Production in Cellulous Fibers (dollars per capita)	IFBR	248
Total Value of Production in Leather and Industrial Leather Products (dollars per capita)	ITAN	248
Total Value of Production in Other Stone and Clay Products (dollars per capita)	IABS	248
Total Value of Production in Copper (dollars per capita)	ICU	248
Total Value of Production in Lead (dollars per capita)	IPB	248
Total Value of Production in Zinc (dollars per capita)	IZN	248
Total Value of Production in Other Fabricated Metal Products (dollars per capita)	IMTL	248
Total Value of Production in Wholesale Trade (dollars per capita)	IWST	248
B. Total Manufacturing		
Total Employment in Manufacturing (percent)	EM	262
Value Added by All Manufacturing (dollars per capita)	VAM	259
C. Industry Mix		
Total Retail Sales ($000 per capita)	RETT	244
Total Wholesale Sales ($000 per capita)	WHOL	249
Selected Services ($000 per capita)	SS	244
Total Value Added in Food and Kindred Products (SIC 20) ($ millions per capita)	S20	246
Total Value Added in Textile Mill Products (SIC 22) ($ millions per capita)	S22	247
Total Value Added in Apparel and Other Textile Mill Products (SIC 23)($ millions per capita)	S23	245
Total Value Added in Lumber and Wood Products (SIC 24) ($ millions per capita)	S24	245
Total Value Added in Furniture and Fixtures (SIC 25) ($ millions per capita)	S25	245

(continued on following page)

Table 3-6 (continued)

	Code	Number of Observations
Total Value Added in Paper and Allied Products (SIC 26)($ millions per capita)	S26	246
Total Value Added in Printing and Publishing (SIC 27)($ millions per capita)	S27	246
Total Value Added in Chemicals and Allied Products (SIC 28)($ millions per capita)	S28	246
Total Value Added in Petroleum and Coal Products (SIC 29) ($ millions per capita)	S29	246
Total Value Added in Rubber and Plastic Products (SIC 30)($ millions per capita)	S30	245
Total Value Added in Leather and Leather Products (SIC 31)($ millions per capita)	S31	245
Total Value Added in Stone, Clay and Glass Products (SIC 32)($ millions per capita)	S32	246
Total Value Added in Primary Metal Industries (SIC 33)($ millions per capita)	S33	244
Total Value Added in Fabricated Metal Products (SIC 34)($ millions per capita)	S34	245
Total Value Added in Machinery, Except Electrical (SIC 35)($ millions per capita)	S35	245
Total Value Added in Electrical Equipment and Supplies (SIC 36)($ millions per capita)	S36	244
Total Value Added in Transportation Equipment (SIC 37)($ millions per capita)	S37	241
Total Value Added in Instruments and Related Products (SIC 38)($ millions per capita)	S38	240
Total Value Added in Miscellaneous Manufacturing Industries (SIC 39)($ millions per capita)	S39	241

SCS2 data, describing urban form and the physical environment, proved to be lacking in descriptors of the distribution of activity within an SMSA. Land use is closely related to environmental quality through the polluting aspects of land uses and through sensitivity of different land uses. Land use data is available for most urban areas. Land use surveys, however, have been performed individually for most SMSAs--the data for each SMSA is collected at a different time by a different agency for a different purpose. Therefore, the data is not consistent in form, nor is it available from a manageable number of sources. Because the effort required to collect sufficient land use data to describe the 262 SMSAs was so great, a number of less reliable and less informative measures were relied upon to describe the distribution of activity within the SMSA.

SCS3, intended to describe the net loading of pollutants on the environment, includes only data from EPA's SEAS data bank for stationary sources. At the time data was collected for this research, reliable estimates of pollution from SEAS for mobile sources in SMSAs were not available,* therefore, the data in this SCS is misleading in terms of the relationships between the actors, or SCSs.

Data available to represent SCS4, households, is extensive since the Census of Population and Housing is concerned with many of the attributes relevant here. Health is an important factor in this SCS and more data on specific diseases known to be related to common pollutants would be desirable and consistent with the research hypotheses. In other respects, however, this data set seems quite satisfactory.

SCS5 is intended to describe the effect of the local government on abatement efforts within the SMSA. Variables, then, are desired to measure government expenditures, regulations, and user charges. Measures of expenditure are readily available. User charges are not widely utilized and, therefore, not quantified on a uniform basis. Regulation of pollution takes on many different forms; for example, air and water emission restrictions, land use zoning, and public transportation subsidies. The best measure of these types of regulation would seem to be a simple indicator of the presence or absence of each type of regulation. Such an indicator would indicate little variation between most SMSAs. Since useful measures of user charges and regulation are not easily obtained, a group of surrogates was used to represent government action in abatement through measures of community participation in the government.

SCS6 is the strongest set of data with respect to the research hypotheses. These data distinguished between SMSAs on the basis of both the magnitude of industrial activity relative to the population and the composition of industrial activity.

Many of the gaps in this data set could be filled in with a substantial effort to collect more data. Volume III documents some sources which may prove helpful in this respect.

3.2.3 Time Frame Represented by the Data

Data utilized for this research is not consistent with respect to time of measurement either as a whole, within each SCS, or even within each variable. In general, this is not a problem given that the factor analysis works with correlations rather than the absolute levels of variables although insight gained from the classification might be greater if all data could be resolved into a single point in time. The time frame covered the period between 1969 and 1975.

---

*The recently implemented SEAS Transportation Energy Demand Module contains estimates of activity levels by SMSA, which may be used to derive residuals estimates.

Most of the data is from measurements taken between 1970 and 1975. Large amounts of data come from the Census of Population and Housing (1970), the Census of Governments (1972), and the Census of Manufacturing (1972). Data comes from many other sources as well, usually for a particular point in time. However, data from STORET, EPA's data bank on water quality, includes individual datums taken by independent sources at different points in time. This means data for a given variable may distort the relative difference in environmental quality between two SMSAs. In reality, the values used for each SMSA from STORET are averages for each parameter over all measurements taken between 1970 and 1975. Therefore, a balance is achieved to some extent even though each parameter might be biased toward a particular time when the frequency of measurement of the given parameter within that SMSA was most intense.

3.2.3 Cities Represented by the Data --
Although 262 cities and approximately 200 variables are represented in the data set, there are missing values within most variables. Measured values were available for a majority of cities on each variable, however, the number of cities for which data is available on all variables is inadequate in some cases. For instance, in SCS1, only 3 variables among the 24 for which data was generated are missing more than 20 percent of the desired observations. Yet the intersection of the data sets is far less complete. Table 3-7 below indicates the number of cases for which data is complete, given the elimination of selected variables.

Table 3-7
Effect of Missing Data on
Sample Size (SCS1)

Number of Variables Utilized	Number of SMSAs with All Data	Percent of SMSAs with All Data
24	53	20
21	91	35
15	122	47
13	134	51
11	137	52

This was a serious problem in SCS1 and led to the eventual use of indicator variables rather than statistically derived factors.

All other SCSs had fewer problems in this respect. Most gaps were associated either with SMSAs in New England (because New England SMSAs are defined according to parts of counties rather than entire counties) or with SMSAs which had been first designated as such between the 1970 Census and 1972, the time selected for this classification. Table 3-8 identifies the numbers of SMSAs in each of these categories.

As the goal of this effort was to classify all SMSAs, rather than those for which every piece of data was available, the data were supplemented with estimates of missing values where a variable was missing data in a small number of cases. Estimates used, in every case, were the statistical means of remaining observations. In no case did estimated values comprise more than 13 percent of observations on a single variable. The use of mean values did not distort the initial factor analysis, which is based on covariances; however, factor scores developed from variable means may not properly represent the SMSAs and can cause distortions in Stage II factor analysis and clustering.

3.2.4 Measurement Problems --

Inconsistencies and inaccuracies in the measurement of variables may be equally as serious a problem as missing observations, missing variables, and temporal inconsistency. The measurement problems referred to here are different for each variable and are documented in Volume III of this report. The most important problems are described here.

3.2.4.1 Ambient Water Quality--Data is retrieved from STORET by identifying geographic areas (in terms of latitudes and longitudes) for which data is desired. The polygon describing each SMSA is necessarily approximate even though as many as 11 points are used to describe a given SMSA boundary. As a result, observations within SMSAs may be excluded in some cases and measurnments outside SMSAs may be included.

Another problem is the uneven distribuiton of measurements over time (as mentioned in Section 3.2.2 above) and over space. STORET data is a collection of data from numerous sources rather than the result of a systematic sampling program. Because the motive for sampling varies, parameters measured, sampling methods, and location type also vary. It is conceivable that measurements from one SMSA might be highly oriented toward recreation facilities while measurements from another SMSA were concentrated at a highly polluted river.

Given measurements of a single variable, BOD for example, for a given time interval and geographic area, STORET provides an estimate of "ambient BOD" by treating all observed measurements as a single station. Although this process yields a usable number, it is inaccurate to assume the available data is representative in terms of variation over time, over space, and between existing water bodies.

3.2.4.2 Ambient Air Quality--The two air quality measures utilized--suspended particulates and sulfur dioxide--do not adequately describe ambient air quality. National Ambient Air Quality Standards have been established for five "criteria pollutants:" suspended particulates, sulfur dioxide, carbon monoxide, oxidants, and nitrogen dioxide. Available data, however, is relatively complete for only the first two parameters. The proportion of SMSAs with missing data for each parameter has been described in Section 3.1.1.

Table 3-8
SMSAs for Which Data is Systematically Missing

Code	Title	Located in New England	Established Between 1970 and 1972	Both
0220	Alexandria, Louisiana		X	
0460	Appleton-Oshkosh, Wisconsin		X	
0780	Battle Creek, Michigan		X	
1120	Boston, Massachusetts	X		
1160	Bridgeport, Connecticut	X		
1170	Bristol, Connecticut	X	X	X
1200	Brockton, Massachusetts	X		
1260	Bryan-College Station, Texas		X	
1740	Columbia, Missouri		X	
1930	Danbury, Connecticut	X	X	X
2020	Daytona Beach, Florida		X	
2335	Elmira, New York		X	
2600	Fitchburg-Leominster, Massachusetts	X		
2650	Florence, Alabama		X	
2700	Fort Myers, Florida		X	
2900	Gainesville, Florida		X	
2970	Gastonia, North Carolina		X	
3280	Hartford, Connecticut	X		
3810	Killeen-Temple, Texas		X	
3870	LaCrosse, Wisconsin		X	
3980	Lakeland-Winter Haven, Florida		X	
4160	Lawrence-Haverhill, Massachusetts-New Hampshire	X		
4240	Lewiston-Auburn, Maine	X		
4410	Long Branch-Asbury Park, New Jersey		X	
4560	Lowell, Massachusetts-New Hampshire	X		
4760	Manchester, New Hampshire	X		
4900	Melbourne-Titusville-Cocoa, Florida		X	
4960	Meriden, Connecticut	X		
5170	Modesto, California		X	
5350	Nashua, New Hampshire	X	X	X
5400	New Bedford, Massachusetts	X		
5440	New Britain, Connecticut	X		
5460	New Brunswick-Perth Amboy-Sayerville, New Jersey		X	
5480	New Haven-West Haven, Connecticut	X		
5760	Norwalk, Connecticut	X		

(continued on following page)

Table 3-8 (continued)

Code	Title	Located in New England	Established Between 1970 and 1972	Both
5990	Owensboro, Kentucky		X	
6020	Parkersburg-Marietta, West Virginia Virginia-Ohio		X	
6140	Petersburg-Colonial Heights-Hopewell, Virginia		X	
6320	Pittsfield, Massachusetts	X		
6400	Portland, Maine	X		
6460	Poughkeepsie, New York		X	
6480	Providence-Warwick-Pawtucket, Rhode Island	X		
6740	Richland-Kennewick, Washington		X	
6820	Rochester, Minnesota		X	
7485	Santa Cruz, California		X	
7500	Santa Rosa, California		X	
7510	Sarasota, Florida		X	
7820	Spartanburg, South Carolina		X	
8000	Springfield-Chicopee-Holyoke, Massachusetts-Connecticut	X		
8040	Stamford, Connecticut	X		
8880	Waterbury, Connecticut	X		
9140	Williamsport, Pennsylvania		X	
9240	Worcester, Massachusetts	X		
9260	Yakima, Washington		X	
	TOTAL	24	33	3

The available data is taken to represent ambient air quality even though measurements are taken at a small number of stations at selected times. Like the water quality data, these tests are not necessarily representative over time and space within an SMSA.

This problem is compounded by the aggregation of raw data. Values for the present research were obtained from a SEAS data file of air quality by air quality control region (AQCR). The 247 AQCRs do not correspond to the 262 SMSAs. AQCR 43, for instance, contains the following SMSAs:

1160 Bridgeport, Connecticut
1930 Danbury, Connecticut
3640 Jersey City, New Jersey
4410 Long Branch-Asbury Park, New Jersey
5460 New Brunswick-Perth Amboy-Sayreville, New Jersey
5600 New York, New York-New Jersey
5640 Newark, New Jersey
5760 Norwalk, Connecticut
6040 Paterson-Clifton-Passaic, New Jersey
8040 Stamford, Connecticut.

Thus, this level of aggregation would suggest Danbury, Connecticut, and New York City have the same air quality.

3.2.4.3 PDI Index (of Ambient Water Quality)--"The PDI index...does not differentiate between the case where a stream reach fails to meet water quality criteria for one pollutant and the case where all pollutant criteria are violated, nor does its value reflect the extent by which a criterion is violated. ...In addition, attention is directed to the fact that certain of the data used in evaluating the indices are in fact estimates made by experts in the Regional Offices rather than physical, chemical, or economic measurements. The numerical values of the index are used primarily for ordering the planning areas and have little significance other than for rank ordering."*

3.2.4.4 Variables Measuring the Proportion of Activity in the Core of an SMSA--The proportions of population, manufacturing, employment, etcetera in the central city(s) of an SMSA were used as measures of central tendency. SMSAs are named in a manner which titles the SMSA by those cities which, by various criteria** appear to act as the core of the SMSA.

There are many reasons why central city data is less than satisfactory in representing the core of an SMSA. First: the criteria for naming SMSAs are arbitrary and do not necessarily reflect the interactions within an SMSA. Second: city boundaries are, in most cases, unrelated to the distribution of activity within an SMSA; therefore, the inclusion of exclusion of an entire city may often be misleading. THird: all data for central cities was taken from the County and City Data Book, 1972 (U.S. Bureau of the Census). Data is reported for cities with populations of 25,000 or more. Fifteen of the SMSAs under study, however, include the names of smaller cities in their titles:

---

*J.B. Truitt, A.C. Johnson, W.D. Row, K.D. Feigner, and L.J. Manning, "Development of Water Quality Management Indices," Water Resources Bulletin (11:3, June 1975).

**See page 19 of Volume III: Data Documentation.

1240	Brownsville-Harlingen-San Benito, Texas
1260	Bryan-College Station, Texas
3920	Lafayette-West Lafayette, Indiana
3980	Lakeland-Winter Haven, Florida
4240	Lewiston-Auburn, Maine
4410	Long Branch-Asbury Park, New Jersey
4880	McAllen-Pharr-Edinburg, Texas
4900	Melbourne-Titusville-Cocoa, Florida
5320	Muskegon-Muskegon Heights, Michigan
6020	Parkersburg-Marietta, West Virginia-Ohio
6140	Petersburg-Colonial Heights-Hopewell, Virginia
6740	Richland-Kennewick, Washington
7640	Sherman-Dennison, Texas
8360	Texarkana, Texas-Texarkana, Arkansas
8760	Vineland-Millville-Bridgeton, New Jersey

Central city data includes only those cities for those SMSAs with populations of 25,000 or more.

3.2.4.5 Residual Pollutants--Data on residuals from the SEAS data bank is computed rather than directly measured. Industry coefficients are developed for approximately 400 pollution producing economic sectors and subsectors. These coefficients relate the generation of a specific pollutant by the particular industry to the output of that industry at the national level.* The coefficient times the output of the industry in the given SMSA equals the total gross residual. A second coefficient estimating abatement by sector is applied to the gross residual at the SMSA level to obtain the net residual.

Industry output at the SMSA level is measured by total value of production. 1975 data used here is actually forecast by the SEAS macromodel rather than measured. The national forecast is shared out between SMSAs based on disaggregate forecasts prepared by the Bureau of Economic Analysis (OBERS), the Economic Information System (EIS) tape, and other appropriate data sources.

Because residuals estimates are based on coefficients for industry sectors, they are insensitive to important sources of variation. First, there is significant variation in industrial processes between regions in the same sector. For example, the Northeast has many very old plants which would be expected to operate very differently from the newer plants in the Midwest. Abatement activities may also vary in a given industrial sector across geographical regions; the SEAS model has some estimates of regional abatement differentials. Further, the composition of industry within each sector will vary between SMSAs. Measuring output as total value of production also produces distortions, since total value of products and amount of pollutant produced are directly related only for a single manufacturing process.

---

*International Research & Technology Corporation, SEAS Phase III Reference Data Documentation, Arlington, VA: 1975.

4.0 STAGE I FACTOR ANALYSIS

As indicated in Chapter 1, the classification of cities took place in three stages:

Stage I: factor analysis of the variables from each SCS separately;

Stage II: factor analysis on all factors resulting from Stage I (combining factors from all SCSs); and

Stage III: geometric clustering on factors resulting from Stage II.

Because the research design placed a great deal of emphasis on conceptual as well as statistical validity at each stage, the procedure and output from each stage is documented separately for this report. Chapters 4, 5, and 6 pertain to Stages I, II and III respectively.

For each SCS, the general approach to developing Stage I factors was to perform initial statistical descriptions and then repeat factor analyses to

1) clear the data set of unrealistic and erroneous observations;

2) modify hypotheses and variable set to reflect knowledge gained from initial analysis;

3) estimate values to replace missing observations;

4) perform final factor analysis.

The sections of this chapter will describe, separately for each SCS, the course of analysis followed and the outcome of Stage I analysis.

4.1 SCS1: Environmental Quality

4.1.1 Missing Data --

The greatest single problem in the analysis of environmental quality data was the number of missing values. As Table 4-1 indicates, no single variable in this set has data values for all cities and most variables are missing a significant number.

The initial factor analysis performed on this set included all the 24 variables listed in Table 4-1. Only 53 cases were includes in the analysis because all others lacked observations on one or more values. The following attempt excluded the 3 variables which contained the fewest observations: Nitrogen (N), Oil and Grease (OAG), and Methylene Blue Active Substance (MBAS), only 91 cases appeared in that analysis. Table 4-1 indicates, though, that with the exception of N, OAG, and MBAS, all variables have values for more than 70 percent of the 262 SMSAs. So, Table 4-2 was generated to identify the variables which are available for similar groups of cities. Reducing the number of variables with the aid of Table 4-2, 15 variables were selected, yielding 122 cases--still less than 50 percent of the 262 SMSAs. At this point it became obvious that missing values for each variable would need to be estimated in order to include more than a small subset of the data available. The method chosen was to replace any missing value with the statistical mean of all other observations on that variable. As Table 4-1 indicates, no more than 20 percent of the observations on any variable are estimates.*

4.1.2 Erroneous Data --

In many cases, data reported for a given SMSA did not appear to be reasonable given expert judgments on environmental quality parameters. Such instances were revealed by unlikely correlations between variables input to the factor analysis and by histograms produced for each variable. Looking at Table 4-3, for instance, it is evident that the observations in the last column are erroneous--perhaps the result of decimal place errors. The remaining 98 percent of observations are all less than 0.6 whereas the last 4 values range from 17 to 300. Observations such as these were treated as missing values and replaced by variable means. A total of 19 reported values were stricken from the data set.

4.1.3 Assimilative Characteristics --

Three variables, originally included in SCS2 (Urban Form and the Physical Environment), have been included in SCS1: dissolved oxygen (DO), Water Temperature (WTMP), Ambient water hardness (HARD). These are among the set of variables which describe the assimilative attributes of the local environment with respect to pollutants. As will be discussed in Section 4.2, all variables measuring assimilative characteristics were eliminated from SCS2 because they tended to dominate those variables describing urban form and seemed to contain information which was captured by factors generated in other SCSs. These 3 variables (DO, WTMP, HARD), in particular, are not really measures of the assimilative capacity as originally hypothesized. Dissolved oxygen, for instance, is a measure of the balance between BOD and the amount of dissolved oxygen in the stream before the addition of a BOD load. As these variables do tend to measure water quality, they have been added to SCS1.

4.1.4 Factor Analysis --

Given the exclusion of inadequately measured variables and unreasonable values, the addition of estimates for missing values, and the addition of 3 new variables, the factor analysis was performed once again to derive Stage I factors. Table 4-4 shows the factors resulting at this point. The first factor indicates heavily polluted and heavily treated (for consumption) water; the second factor indicates a warm climate; the third describes an alkaline water environment; the fourth indicates heavy organic loads in the water; etcetera. The 9 factors shown are those with eigenvalues greater than 1; in other words, they explain more than a random amount of the total variance contained in the original data. Together they explain 66 percent of the variance in the original data set. With one exception, the variables have good communalities: TDSS has an extremely low communality of 0.29.** Statistically, then, the factors provide adequate surrogates for the original variables.

---

*N, OAG, and MBAS are not included in the factor analysis.

**The communality of a variable is its squared multiple correlation with the factors.

Table 4-1

ECSI Missing Variables

Variable	Missing Values	Percent of Observations Missing			
		0-10%	11-20%	21-30%	31-40%
Biochemical Oxygen Demand	31		X		
Fecal Coliforms	49		X		
Total Nitrogen	90				X
Total Organic Carbon	57			X	
Dissolved Solids	37		X		
Suspended Solids	47		X		
Turbidity	34		X		
Alkalinity	11	X			
Acidity	7	X			
Oil and Grease	113				X
Methylene Blue Active Substance	101			X	
Suspended Particulates	6	X			
Sulfur Dioxide	29		X		
PDI Index	20	X			
Chloride	55			X	
Fluoride	60			X	
Iron	56			X	
Manganese	61			X	
Nitrate	66			X	
Sulfate	70			X	
Alkalinity	73			X	
Hardness	71			X	
Acidity	58			X	
Total Dissolved Solids	74			X	

Table 4-1

Environmental Quality:
Missing Data

SMSA	AMBIENT WATER QUALITY									AMBIENT AIR QUALITY	
	STREAM								Prevalence, Duration & Intensity Index		
	pH	BOD	Fecal Coliforms	Carbon	Total Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity		Suspended Particulates	Sulfur Dioxide
0040 Abilene, TX		✓	✓	✓	✓	✓	✓				
0080 Akron, OH											
0120 Albany, GA			✓								
0160 Albany-Schen.-Troy, NY											✓
0200 Albuquerque, NM									✓		
0220 Alexandria, LA											
0240 Allentown-Beth.-East.-PA.-NJ											✓
0280 Altoona, PA											
0320 Amarillo, TX			✓								
0360 Anaheim-San. Ana-Gar. Grove, CA				✓			✓				
0400 Anderson, IN		✓		✓	✓						
0440 Ann Arbor, MI											
0460 Appleton-Osh., WI											
0480 Asheville, NC		✓		✓							
0520 Atlanta, GA											✓
0560 Atlantic City, NJ											
0600 Augusta, GA-SC											
0640 Austin, TX											
0680 Bakersfield, CA	✓	✓	✓	✓	✓	✓	✓	✓	✓		
0720 Baltimore, MD											
0760 Baton Rouge, LA									✓		
0780 Battle Creek, MI											
0800 Bay City, MI											
0840 Beaumont-Port Art-Orange, TX		✓	✓	✓	✓	✓	✓			✓	✓
0880 Billings, MT											
0920 Biloxi-Gulfpt., MS											✓
0960 Binghamton, NY-PA											
1000 Birmingham, AL			✓				✓			✓	✓
1040 Bloomington-Nor., IL											✓
1080 Boise City, ID											
1120 Boston, MA											✓
1160 Bridgeport, CT	✓	✓	✓	✓	✓	✓	✓	✓			
1170 Bristol, CT				✓		✓					
1200 Brockton, MA											
1240 Brownsville-Har.San Ben.-Harl., TX											✓
1260 Bryan-Coll. Sta., TX		✓	✓	✓	✓	✓	✓				
1280 Buffalo, NY											
1320 Canton, OH											
1360 Cedar Rapids, IA				✓			✓			✓	✓
1400 Champaign-Urb.-Rant., IL											
1440 Charleston, SC											
1480 Charleston, WV											
1520 Charlotte, NC											
1560 Chattanooga, TN-GA											
1600 Chicago, IL											
1640 Cincinnati, OH-KY-IN											

a/ [illegible]

(continued on following page)

Table 4-2: continued

	DRINKING WATER QUALITY									
SMSA	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
0040 Abilene, TX										
0080 Akron, OH										
0120 Albany, GA										
0160 Albany-Schen.-Troy, NY										
0200 Albuquerque, NM										
0220 Alexandria, LA				a/						
0240 Allentown-Beth.-East., PA-NJ										
0280 Altoona, PA										
0320 Amarillo, TX								a/		
0360 Anaheim-San.Ana-Gar. Grove, CA	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0400 Anderson, IN	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0440 Ann Arbor, MI										a/
0460 Appleton-Osh., WI	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0480 Asheville, NC	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0520 Atlanta, GA										
0560 Atlantic City, NJ										
0600 Augusta, GA-SC										
0640 Austin, TX										
0680 Bakersfield, CA										
0720 Baltimore, MD										
0760 Baton Rouge, LA										
0780 Battle Creek, MI	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0800 Bay City, MI	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0840 Beaumont-Port Arth-Orange, TX										
0880 Billings, MT							a/	a/		
0920 Biloxi-Gulfpt., MS	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
0960 Binghamton, NY-PA		a/				a/				a/
1000 Birmingham, AL										a/
1040 Bloomington-Nor., IL	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
1080 Boise City, ID							a/	a/	a/	
1120 Boston, MA										
1160 Bridgeport, CT						a/				a/
1170 Bristol, CT	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
1200 Brockton, MA	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
1240 Brownsville-Har.-San Ben., TX										
1260 Bryan-Coll. Sta., TX	a/	a/	a/	a/	a/	a/	a/	a/	a/	a/
1280 Buffalo, NY										
1320 Canton, OH										
1360 Cedar Rapids, IA										
1400 Champaign-Urb.-Rant., IL					a/					
1440 Charleston, SC		a/			a/	a/				
1480 Charleston, WV										
1520 Charlotte, NC										
1560 Chattanooga, TN-GA										
1600 Chicago, IL										
1640 Cincinnati, OH-KY-IN	a/									

a/ - data missing

(continued on following page)

Table 4-1: continued

SMSA	AMBIENT WATER QUALITY									AMBIENT AIR QUALITY	
	STORET								Prevalence, Duration & Intensity Index		
	pH	BOD	Fecal Coliforms	Carbon	Solids Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity		Suspended Particulates	Sulfur Dioxide
1680 Cleveland, OH											
1720 Colorado Springs, CO											✓
1740 Columbia, MO				✓		✓		✓			✓
1760 Columbia, SC											
1800 Columbus, GA-AL			✓			✓					
1840 Columbus, OH											
1880 Corpus Christi, TX			✓			✓					
1920 Dallas, TX			✓								
1930 Danbury, CT		✓	✓	✓	✓	✓	✓				
1960 Davenport-Rock Is.-Mol., IA-IL											
2000 Dayton, OH											
2020 Daytona Beach, FL									✓		✓
2040 Decatur, IL				✓			✓				
2080 Denver-Boulder, CO											
2120 Des Moines, IA											
2160 Detroit, MI											
2200 Dubuque, IA				✓							
2240 Duluth-Sup., MN-WI											
2280 Durham, NC				✓							
2320 El Paso, TX											
2335 Elmira, NY			✓	✓		✓			✓		
2360 Erie, PA											
2400 Eugene-Spr'fld., OR				✓							
2440 Evansville, IN-KY											
2480 Fall River, MA-RI			✓	✓							
2520 Fargo-Moor., ND-MN											
2560 Fayetteville, NC			✓	✓							
2600 Fitchburg, Leom., MA				✓							
2640 Flint, MI											
2650 Florence, AL									✓		
2680 Ft. Lauderdale-Holly., FL											
2700 Fort Myers, FL									✓	✓	✓
2720 Fort Smith, AR-OK											
2760 Fort Wayne, IN						✓					
2800 Fort Worth, TX			✓	✓	✓		✓				
2840 Fresno, CA		✓									
2880 Gadsden, AL				✓							✓
2900 Gainesville, FL			✓								
2920 Galveston-Tx.City, TX				✓							
2960 Gary-Ham.-E.Chi., IN											
2975 Gastonia, NC									✓		
3000 Grand Rapids, MI											
3040 Great Falls, MT			✓	✓			✓				
3080 Green Bay, WI											
3120 Greensboro-Wins.-Sal.-High Pt., NC											

✓- data missing

(continued on following page)

Table 4-1. continued

SMSA	DRINKING WATER QUALITY									
	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
1680 Cleveland, OH										
1720 Colorado Springs, CO							√	√		
1740 Columbia, MO							√	√		
1760 Columbia, SC							√	√		
1800 Columbus, GA-AL										
1840 Columbus, OH										
1880 Corpus Christi, TX										
1920 Dallas, TX										
1930 Danbury, CT										
1960 Davenport-Rock Is.-Mol., IA-IL						√				√
2000 Dayton, OH										
2020 Daytona Beach, FL					√					
2040 Decatur, IL	√	√	√	√	√	√	√	√	√	√
2080 Denver-Boulder, CO										
2120 Des Moines, IA										
2160 Detroit, MI										
2200 Dubuque, IA										√
2240 Duluth-Sup., MN-WI							√	√		
2280 Durham, NC					√					
2320 El Paso, TX										
2335 Elmira, NY	√	√	√	√	√	√	√	√	√	√
2360 Erie, PA										
2400 Eugene-Spfld., OR										
2440 Evansville, IN-KY										
2480 Fall River, MA-RI										
2520 Fargo-Moor., ND-MN										
2560 Fayetteville, NC					√	√				√
2600 Fitchburg, Leom., MA										
2640 Flint, MI	√	√	√	√	√	√	√	√	√	√
2650 Florence, AL										√
2680 Ft. Lauderdale-Holly., FL										
2700 Fort Myers, FL							√	√	√	
2720 Fort Smith, AR-OK		√								
2760 Fort Wayne, IN	√	√	√	√	√	√	√	√	√	√
2800 Fort Worth, TX										
2840 Fresno, CA										
2880 Gadsden, AL	√	√	√	√	√	√	√	√	√	√
2900 Gainesville, FL	√	√	√	√	√	√	√	√	√	√
2920 Galveston-Tx.City, TX										
2960 Gary-Ham.-E.Chi., IN										
[illegible] Gastonia, NC	√	√	√	√	√	√	√	√	√	√
3000 Grand Rapids, MI										
3040 Great Falls, MT										
3080 Green Bay, WI							√	√		
3120 Greensboro-Wins.-Sal.-High Pt., NC							√	√		

√ = data missing

continued on following page

Table 6-1 continued

SMSA	AMBIENT WATER QUALITY									AMBIENT AIR QUALITY	
	STREAM								Prevalence, Duration & Intensity Index		
	pH	BOD	Fecal Coliforms	Carbon	Total Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity		Suspended Particulates	Sulfur Dioxide
3160 Greenville-Spar., SC											
3200 Hamilton-Middle., OH											
3240 Harrisburg, PA											
3280 Hartford, CT											
3360 Houston, TX											
3400 Huntington-Ash., WV-KY-OH											
3440 Huntsville, AL				✓							
3480 Indianapolis, IN											
3520 Jackson, MI											
3560 Jackson, MS											
3600 Jacksonville, FL											
3640 Jersey City, NJ						✓					
3680 Johnstown, PA											
3720 Kalamazoo-Port., MI											✓
3760 Kansas City, MO-KS											
3800 Kenosha, WI				✓		✓	✓				
3810 Killeen-Temple, TX		✓	✓	✓	✓	✓	✓		✓		
3840 Knoxville, TN											
3870 La Crosse, WI											
3880 Lafayette, LA		✓	✓	✓							
3920 Lafayette-W. Lafay., IN				✓							
3960 Lake Charles, LA											
3980 Lakeland-Win. Hav., FL											
4000 Lancaster, PA									✓		
4040 Lansing-E. Lans., MI											
4080 Laredo, TX											
4120 Las Vegas, NV											✓
4160 Lawrence-Haver., MA-NH											
4200 Lawton, OK		✓	✓	✓	✓		✓				
4240 Lewiston-Aub., ME											
4280 Lexington, KY			✓	✓	✓		✓				
4320 Lima, OH											
4360 Lincoln, NE										✓	✓
4400 Little Rock-N.Lit.Rock, AR											
4410 Long Branch-Asb. Pk, NJ											
4440 Lorain-Elyria, OH									✓		
4480 Los Ang-Lon-Lo. Beach, CA											
4520 Louisville, KY-IN											
4560 Lowell, MA-NH											
4600 Lubbock TX		✓	✓	✓	✓	✓	✓				
4640 Lynchburg, VA											
4680 Macon, GA			✓								
4720 Madison, WI				✓							
4760 Manchester, NH						✓					
4800 Mansfield, OH		✓	✓		✓						
4880 McAllen-Pharr-Edinb., TX		✓	✓	✓	✓	✓	✓				✓
4900 Melbourne-Titusvl.-Cocoa, FL											
4920 Memphis, TN-AR-MS									✓		
4960 Meriden, CT	✓	✓	✓	✓	✓	✓	✓	✓			

✓ = data missing

(continued on following page)

Table 4-2: continued

SMSA	DRINKING WATER QUALITY									
	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
3160 Greenville-Spar., SC					✓	✓				
3200 Hamilton-Middle., OH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3240 Harrisburg, PA										
3280 Hartford, CT										✓
3360 Houston, TX										
3400 Huntington-Ash., WV-KY-OH										
3440 Huntsville, AL							✓	✓		
3480 Indianapolis, IN										
3520 Jackson, MI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3560 Jackson, MS							✓	✓		
3600 Jacksonville, FL										
3640 Jersey City, NJ										
3680 Johnstown, PA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3720 Kalamazoo-Port., MI										
3760 Kansas City, MO-KS										
3800 Kenosha, WI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3810 Killeen-Temple, TX										
3840 Knoxville, TN										
3870 La Crosse, WI							✓	✓		
3880 Lafayette, LA										
3920 Lafayette-W. Lafay., IN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3960 Lake Charles, LA										
3980 Lakeland-Win. Hav., FL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4000 Lancaster, PA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4040 Lansing-E. Lans., MI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4080 Laredo, TX										
4120 Las Vegas, NV				✓						
4160 Lawrence-Haver., MA-NH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4200 Lawton, OK		✓								
4240 Lewiston-Aub., ME		✓				✓	✓			✓
4280 Lexington, KY										
4320 Lima, OH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4360 Lincoln, NE										
4400 Little Rock-N. Lit. Rock, AR										
4420 Long Branch-Asb. Pk, NJ										
4440 Lorain-Elyria, OH										
4480 Los Angeles-Lg. Beach, CA										
4520 Louisville, KY-IN										
4560 Lowell, MA-NH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4600 Lubbock TX										
4640 Lynchburg, VA										
4680 Macon, GA										
4720 Madison, WI							✓	✓		
4760 Manchester, NH						✓				✓
4800 Mansfield, OH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
4880 McAllen-Phar.-Edinb., TX										
4890 Melbourne-Titusvl.-Cocoa, FL										
4920 Memphis, TN-AR-MS										
4960 Meriden, CT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

✓ = Data missing

(continued on following page)

Table 4-2: continued

SMSA	AMBIENT WATER QUALITY									AMBIENT AIR QUALITY	
	STREAMS								Occurrence, Duration & Intensity Index		
	pH	BOD	Fecal Coliforms	Carbon	Total Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity		Suspended Particulates	Sulfur Dioxide
5000 Miami, FL											
5040 Midland, TX	✓	✓	✓	✓	✓	✓	✓	✓			
5080 Milwaukee, WI											
5120 Minneapolis-St. Paul, MN-WI											
5160 Mobile, AL			✓								
5170 Modesto, CA		✓									
5200 Monroe, LA											
5240 Montgomery, AL											
5280 Muncie, IN			✓	✓		✓					
5320 Muskegon-Musk. Hts, MI			✓			✓					
5350 Nashua, NH											
5360 Nashville-Davidson, TN											
5400 New Bedford, MA		✓	✓	✓	✓	✓	✓				
5440 New Britain, CT	✓	✓	✓	✓	✓	✓	✓	✓			
5460 New Brunswick-Perth Amboy-Sayreville, NJ						✓		✓	✓		
5480 New Haven-W. Haven, CT											
5520 New London-Norw., CT-RI											✓
5560 New Orleans, LA											
5600 New York, NY, NJ											
5640 Newark, NJ											
5680 Newport News-Hamp., VA					✓	✓	✓	✓			
5720 Norfolk-Va. Beach-Ports., VA-NC					✓						
5760 Norwalk, CT	✓	✓	✓	✓	✓	✓		✓			
5800 Odessa, TX	✓	✓	✓	✓	✓	✓	✓	✓			
5840 Ogden, UT											
5880 Oklahoma City, OK											
5920 Omaha, NE-IA											
5960 Orlando, FL											
5990 Owensboro, KY											
6000 Oxnard-Simi Val.-Ven., CA			✓		✓						
6020 Parkersburg-Marie., WV-OH									✓		✓
6040 Paterson-Clif-Passa., NJ											
6080 Pensacola, FL											
6120 Peoria, IL						✓	✓				
6140 Petersburg-Col. Hts-Hope., VA			✓		✓	✓					
6160 Philadelphia, PA-NJ											
6200 Phoenix, AZ											
6240 Pine Bluff, AR								✓			
6280 Pittsburgh, PA											
6320 Pittsfield, MA		✓	✓	✓	✓	✓	✓				
6400 Portland, ME											
6440 Portland, OR-WA											
6460 Poughkeepsie, NY									✓		
6480 Providence-War.-Paw., RI-MA											
6520 Provo-Orem, UT				✓							
6560 Pueblo, CO				✓							
6600 Racine, WI		✓	✓	✓	✓	✓	✓				
6640 Raleigh, NC											
6680 Reading, PA											

Table 4-1. continued

SMSA	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
	DRINKING WATER QUALITY									
5000 Miami, FL										
5040 Midland, TX										
5080 Milwaukee, WI							✓	✓		
5120 Minneapolis-St. Paul, MN-WI										
5160 Mobile, AL							✓	✓		
5170 Modesto, CA										
5200 Monroe, LA										
5240 Montgomery, AL										✓
5280 Muncie, IN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5320 Muskegon-Musk. Hts., MI										
5350 Nashua, NH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5360 Nashville-Davidson, TN					✓					
5400 New Bedford, MA										
5440 New Britain, CT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5460 New Brunswick-Perth Amboy-Sayreville, NJ										
5480 New Haven-W. Haven, CT						✓				✓
5520 New London-Norw., CT-RI						✓				✓
5560 New Orleans, LA										
5600 New York, NY, NJ										
5640 Newark, NJ										
5680 Newport News-Hamp., VA										
5720 Norfolk-Va. Beach-Ports., VA-NC										
5760 Norwalk, CT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5800 Odessa, TX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
5840 Ogden, UT										
5880 Oklahoma City, OK				✓	✓					
5920 Omaha, NE-IA										
5960 Orlando, FL										
5990 Owensboro, KY	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6000 Oxnard-Sim. Val.-Ven., CA										
6020 Parkersburg-Marie., WV-OH										
6040 Paterson-Clif-Pass., NJ										
6080 Pensacola, FL							✓	✓		
6120 Peoria, IL										
6140 Petersburg-Col. Hts-Hope., VA										
6160 Philadelphia, PA-NJ										
6200 Phoenix, AZ				✓						
6240 Pine Bluff, AR										✓
6280 Pittsburgh, PA										
6320 Pittsfield, MA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6400 Portland, ME						✓				
6440 Portland, OR-WA										
6460 Poughkeepsie, NY		✓				✓				✓
6480 Providence-War.-Paw., RI-MA										
6520 Provo-Orem, UT	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6560 Pueblo, CO									✓	
6600 Racine, WI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6640 Raleigh, NC					✓					✓
6680 Reading, PA										

✓ = [illegible]

continued on following page

Table 4-2: continued

SMSA	AMBIENT WATER QUALITY STORET pH	BOD	Fecal Coliform	Carbon	Total Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity	Prevalence, Duration & Intensity Index	AMBIENT AIR QUALITY Suspended Particulates	Sulfur Dioxide
6720 Reno, NV					✓						✓
6740 Richland-Kenne., WA									✓		
6760 Richmond, VA											
6780 Riverside-San Bern.-Ont., CA											
6800 Roanoke, VA											
6820 Rochester, MN											
6840 Rochester, NY											
6880 Rockford, IL		✓		✓	✓	✓	✓				
6920 Sacramento, CA											
6960 Saginaw, MI											
6980 St. Cloud, MN											
7000 St. Joseph, MO											
7040 St. Louis, MO-IL											
7080 Salem, OR											
7120 Salinas-Seas.-Mont., CA					✓						
7160 Salt Lake City, UT											
7200 San Angelo, TX		✓	✓	✓	✓	✓	✓				
7240 San Antonio, TX			✓								
7320 San Diego, CA		✓	✓	✓	✓		✓				
7360 San Francisco-Oak., CA											
7400 San Jose, CA					✓		✓				
7480 Santa Barbara-San.Maria-Lompoc, CA		✓	✓	✓							✓
7485 Santa Cruz, CA				✓	✓				✓		✓
7500 Santa Rosa, CA		✓			✓						
7510 Sarasota, FL									✓	✓	✓
7520 Savannah, GA											
7560 Scranton, PA											
7600 Seattle-Everett, WA											
7640 Sherman-Denis., TX											
7680 Shreveport, LA											
7720 Sioux City, IA-NE											
7760 Sioux Falls, SD											
7800 South Bend, IN			✓	✓		✓					
7820 Spartanburg, SC									✓		
7840 Spokane, WA			✓								
7880 Springfield, IL											✓
7920 Springfield, MO				✓		✓	✓				
7960 Springfield, OH		✓									
8000 Springfield-Chico.-Holy., MA-CT						✓					
8040 Stamford, CT											
8080 Steubenville-Weirt.,OH-WV		✓	✓	✓							
8120 Stockton, CA											
8160 Syracuse, NY											

✓- data missing

(continued on following page)

Table 4-1. continued

SMSA	DRINKING WATER QUALITY									
	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
6720 Reno, NV				√						
6740 Richland-Kenne., WA										
6760 Richmond, VA										
6780 Riverside-San Bern.-Ont., CA	√	√	√	√	√	√	√	√	√	√
6800 Roanoke, VA										
6820 Rochester, MN	√	√	√	√	√	√	√	√	√	√
6840 Rochester, NY					√					
6880 Rockford, IL	√	√	√	√	√	√	√	√	√	√
6920 Sacramento, CA										
6960 Saginaw, MI										
6980 St. Cloud, MN										
7000 St. Joseph, MO	√	√	√	√	√	√	√	√	√	√
7040 St. Louis, MO-IL										
7080 Salem, OR	√	√	√	√	√	√	√	√	√	√
7120 Salinas-Seas.-Mont., CA								√		
7160 Salt Lake City, UT										
7200 San Angelo, TX										
7240 San Antonio, TX										
7320 San Diego, CA										
7360 San Francisco-Oakl., CA										
7400 San Jose, CA										
7480 Santa Barbara-San.Maria-Lompoc, CA										
7485 Santa Cruz, CA	√	√	√	√	√	√	√	√	√	√
7500 Santa Rosa, CA	√	√	√	√	√	√	√	√	√	√
7510 Sarasota, FL			√	√		√	√	√	√	
7520 Savannah, GA										
7560 Scranton, PA										
7600 Seattle-Everett, WA										
7640 Sherman-Denis., TX	√	√	√	√	√	√	√	√	√	√
7680 Shreveport, LA										
7720 Sioux City, IA-NE	√	√	√	√	√	√	√	√	√	√
7760 Sioux Falls, SD										
7800 South Bend, IN										
7820 Spartanburg, SC										
7840 Spokane, WA										
7880 Springfield, IL										
7920 Springfield, MO										
7960 Springfield, OH	√	√	√	√	√	√	√	√	√	√
8003 Springfield-Chicop.-Holy., MA-CT										
8040 Stamford, CT						√				√
8080 Steubenville-Weirt., OH-WV										
8120 Stockton, CA										
8160 Syracuse, NY										

√ = data missing

(continued on following page)

Table 4-2: continued

SMSA	AMBIENT WATER QUALITY									AMBIENT AIR QUALITY	
	STORET								Prevalence-Duration-Intensity Index		
	pH	DO	Fecal Coliforms	Carbon	Total Dissolved Solids	Total Suspended Solids	Turbidity	Alkalinity		Suspended Particulates	Sulfur Dioxide
[illegible] Tacoma, WA											
[illegible] Tallahassee, FL											
[illegible] Tampa-St.Petersburg, FL											
[illegible] Terre Haute, IN											
[illegible] Texarkana, TX-Texarkana, AR											
[illegible] Toledo, OH-MI											
[illegible] Topeka, KS											
[illegible] Trenton, NJ											
[illegible] Tucson, AZ				✓		✓					
[illegible] Tulsa, OK											
[illegible] Tuscaloosa, AL											
[illegible] Tyler, TX			✓			✓					
[illegible] Utica-Rome, NY											
[illegible] Vallejo-Fairfld.-Napa, CA											
[illegible] Vineland-Millvl.-Bridgeton, NJ			✓								✓
[illegible] Waco, TX			✓	✓	✓	✓	✓				
[illegible] Washington, DC-MD-VA											
[illegible] Waterbury, CT					✓						
[illegible] Waterloo-Cedar Falls, IA											
[illegible] West Palm Beach-Boc.Raton, FL											
[illegible] Wheeling, WV-OH			✓								
[illegible] Wichita, KS											
[illegible] Wichita Falls, TX					✓						
[illegible] Wilkes Barre-Hazl., PA											
[illegible] Williamsport, PA									✓		✓
[illegible] Wilmington, DE-NJ-MD											
[illegible] Wilmington, NC				✓							
[illegible] Worcester, MA											
[illegible] Yakima, WA											
[illegible] York, PA											
[illegible] Youngstown-Warren, OH											

✓ - data missing

(continued on following page)

Table 4-1: continued

	DRINKING WATER QUALITY									
	Chloride	Fluoride	Iron	Manganese	Nitrate	Sulfate	Alkalinity	Hardness	pH	Total Dissolved Solids
[illegible] Tacoma, WA										
[illegible] Tallahassee, FL										
[illegible] Tampa-St. Petersburg, FL										
[illegible] Terre Haute, IN	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[illegible] Texarkana, TX- Texarkana, AR										
[illegible] Toledo, OH-MI										
[illegible] Topeka, KS										
[illegible] Trenton, NJ										
[illegible] Tucson, AZ										
[illegible] Tulsa, OK										
[illegible] Tuscaloosa, AL										
[illegible] Tyler, TX	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[illegible] Utica-Rome, NY										
[illegible] Vallejo-Fairfield-Napa, CA										
[illegible] Vineland-Millville- Bridgeton, NJ							✓			
[illegible] Waco, TX										
[illegible] Washington, DC-MD-VA										
[illegible] Waterbury, CT										
[illegible] Waterloo-Cedar Falls, IA						✓				✓
[illegible] West Palm Beach-Boca Raton, FL										
[illegible] Wheeling, WV-OH	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[illegible] Wichita, KS										
[illegible] Wichita Falls, TX										
[illegible] Wilkes-Barre-Hazleton, PA										
[illegible] Williamsport, PA										
[illegible] Wilmington, DE-NJ-MD										
[illegible] Wilmington, NC										
[illegible] Worcester, MA										
[illegible] Yakima, WA										
[illegible] York, PA	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[illegible] Youngstown-Warren, OH										✓

a/ [illegible] missing

```
VARIABLE NAME . . . . . . . [illegible]        [illegible]   [illegible]                                  [illegible]
         NAME . . . . . . . [illegible]        [illegible]   [illegible]
NUMBER OF DISTINCT VALUES .  [illegible]       RANGE         [illegible]
NUMBER OF VALUES COUNTED. .  [illegible]       VARIANCE      [illegible]                                  EACH MARK
NUMBER OF VALUES NOT COUNTED [illegible]       ST.DEV.       [illegible]                                  REPRESENTS
                                               [illegible]   [illegible]                                    [illegible]
LOCATION ESTIMATES                                   ST.ERROR                                                COUNTS
                 MEAN         [illegible]    [illegible]
                 MEDIAN       [illegible]    [illegible]
                 MODE         [illegible]

                                                                                             EACH . =   [illegible]

        PERCENTS                       PERCENTS                        PERCENTS                       PERCENTS
VALUE COUNT CELL  CUM    VALUE  COUNT  CELL  CUM     VALUE   COUNT   CELL   CUM      VALUE  COUNT  CELL   CUM
[illegible]                  [illegible]                      0.17     1      0.5   95.5     17.33    1     0.5   98.5
                                                         [illegible]  1      0.5   96.0     [illegible]  1  0.5   99.0
                                                         [illegible]  1      0.5   96.5     [illegible]  1  0.5   99.5
                                                         [illegible]  1      0.5   97.0     300.00   1     0.5  100.0
                                                             0.21     1      0.5   97.5
                                                         [illegible]  1      0.5   98.0
```

Table 4-3

Distribution of Manganese Values Observed

Table 4-4

Stage I Factor Analysis - SYSI

(All Variables except %, CAG, MBAS)

Variable	Factor								
	1	2	3	4	5	6	7	8	9
Total Dissolved Solids	++								
Sulfate	++								
Hardness	+								
Chloride	+								
Suspended Particulates	+								
Alkalinity (Drinking Water Quality)	+						+		
Water Temperature		++							
Dissolved Oxygen		--							
Acidity (Drinking Water Quality)			+						
Acidity (Ambient Water Quality)			+						
Alkalinity (Ambient Water Quality)			+	++					
Biochemical Oxygen Demand				++					
Fecal Coliform				+					
Total Organic Carbon					++				
Suspended Solids					+				
Hardness						++			
Iron						++			
Fluoride							++		
Nitrate								++	
Sulfur Dioxide									++
PSI Index									
Manganese									
Turbidity								+*	
Dissolved Solids									

*This loading is less than .50.

Factor Loadings: .75 to 1.00 (++)
.50 to .75 (+)
-.50 to -.75 (-)
-.75 to -1.00 (--)

The character of the original data set, however, indicates that these factors may not be the best available descriptors of environmental quality. One reason is that the variables included in this data set emphasize ambient water quality and drinking water quality. The number of variables in each area is shown in Table 4-5.

Table 4-5

Distribution of Environmental Quality Variables

Ambient Water Quality	11
Ambient Air Quality	2
Subjective Indices	1
Drinking Water Quality	10

Factor analysis on these variables will yield factors which describe water quality almost exclusively. Another problem with the data set is the large number of estimated values which replace missing values. Although a series with estimated values included will generally add information to the factor analysis, each estimated value makes the series less reliable.

The foregoing factor analysis seems valid, however, there is some question as to the reliability of the original data set. Therefore, the decision was made to use actual indicator variables rather than factors as the output from Stage I. Indicator variables were chosen using three criteria:

- degree to which the variables differentiate between different SMSAs on the basis of environmental quality;
- number of observations missing from the data series;
- degree to which the variables represent the different factors generated from the set of 24 variables.

The following variables were selected: Suspended Particulates (SUS), Dissolved Oxygen (DO), Acidity (PHS), Hardness (HARD), Sulfur Dioxide (SO2), Index of Water Quality (PDI), and Turbidity (TURB).[*] Table 4-6 shows the relationships of these variables to the criteria enumerated above. Variables representing drinking water quality were eliminated altogether; as a whole, this set had a large proportion of missing data. In addition, measurements of drinking water were found to be inaccurate.

---

[*] See Table 3-1.

Table 4-6

Indicator Variables

	CODE	NAME	FACTOR	OBSERVATIONS
Ambient Water Quality	DO	Dissolved Oxygen	2	246
	PHS	Acidity	3	255
	BOD	Biochemical Oxygen Demand	4	231
	HARD	Hardness	5	249
	TURB	Turbidity	6	230
Ambient Air Quality	SUS	Suspended Particulates	1	256
	SO2	Sulfur Dioxide	8	233
Subjective Indices	PDI	Index of Water Quality	9	242

A factor analysis performed with these 8 variables indicates they were highly independent and could be treated as factors.

4.2 SCS2: Urban Form and the Physical Environment

Analysis of SCS2 included a large number of factor analysis runs--each one yielding new information about the variable measures involved. Form the 30 variables originally included in this SCS* 15 were chosen for the final Stage I factors which best reflected the urban form concept of the original research hypothesis.

The initial step was the elimination of 2 variables which had far too few observations to be useful:

mean annual possible sunlight	136 observations
mean wind speed	191 observations.

The remaining 28 variables had at least 220 observations, and mean values of the variables were used to estimate the missing observation.**

4.2.1 Initial Factor Analysis--

Initial factor analysis of these 28 variables yielded 10 factors which together explained 68.8 percent of the variance in the original data set. Of these, the first 5 (those which explain the highest proportion of that variance) are portrayed in Table 4-7. Note the 5 factors shown do not include all variables--10 of the more interesting variables load heavily only in the 4 last factors. The first 5 factors have been interpreted as follows:

---

*See Table 3-1.

**Only 120 SMSAs had data for all 28 variables.

1. Region: east-west for the most part

2. Concentration of population and service activities in the central city (cities)

3. Region: north-south

4. Concentration of employment and manufacturing in the central city (cities)

5. Uninterpretable.

This set of factors is very clearly dominated by regional indicators. Clustering on the 9 factor scores yielded groups such as the following:

0040	Abilene, Texas
5040	Midland, Texas
7200	San Angelo, Texas
0320	Amarillo, Texas
4600	Lubbock, Texas
5990	Owensboro, Kentucky
7240	San Antonio, Texas
0640	Austin, Texas
5800	Odessa, Texas
1880	Corpus Christi, Texas

and

7500	Santa Rosa, California
7400	San Jose, California
6920	Sacramento, California
8120	Stockton, California
7320	San Diego, California
7485	Santa Cruz, California

Although regional influences are important to environmental quality, regional effects are evident in other SCEs as well. Therefore, allowing this SCE to be dominated by regional factors would overemphasize regional effects--perhaps to the point of excluding all others. In this SCE in particular, variation in urban form has been made secondary to regional factors.

4.2.2 Elimination of Selected Variables --
In order to reassert the hypothesis that urban form is important to environmental quality, all measures of assimilative attributes were removed from this SCE. In addition, 2 other variables were eliminated: land in farms (LA) and vehicle-miles per capita (VMPP).

---

*The number of water bodies should be inversely rather than directly related to the number of quarters of the metropolitan area. No explanation is available for the high correlation between water hardness and these other two variables.

Table 4-7

BCS2 - Stage I - Initial Factor Analysis

Variable	Factor 1	2	3	4	5
Longitude of SMSA	++				
Mean Annual Precipitation	-				
Inversions (mean annual)	+				
Square miles Per Person	+				
Altitude	+				
Portion of Land Which is Urbanized	-				
Portion of Total Population in Central City		++			
Portion of Retail Sales in Central City		++			
Portion of Manufacturing in Central City		++			
Land in Farms		+			
Water Temperature (ambient)			--		
Latitude of SMSA			++		
Dissolved Oxygen			++		
Portion of Manufacturing Employment in Central City				++	
Portion of Employment in Central City				++	
Large Water Bodies					+
Arc of SMSA around Center					+
Hardness of Water (ambient)					-
Total Miles of Roadway per Capita					
Portion of Principal Arterials					
Number of Major Circumferential Roadways					
Number of Major Radial Roadways					
Portion of Land Devoted to Outdoor Recreation					
Number of Population Centers					
Square Miles Per Person -- Urban Places					
Vehicle Miles Travelled per Capita					
Portion of Workers Working Outside SMSA					
Square Miles Per Person -- Central City					

*Factor loadings are indicates as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--).

Land in farms is taken from the 1969 Census of Agriculture and is available by county only. Because all but New England SMSAs are comprised of entire counties, this variable tends to separate New England SMSAs from all others. Although this was not obvious from the factors in Table 4-7, the effect appeared (from correlation coefficients) to be sufficiently strong to distort the underlying characteristics of cities.

Total vehicle miles per capita appears not to be correlated with other indicators of urban form. It does not load on the major factors and the highest correlation with another variable is .176 with urban land (LU); correlations with all other variables are much lower. The lack of correlation is probably because VMPP is estimated from gasoline sales and road mileage; the former is available at the State level rather than the metropolitan level, thus, an accurate estimate of VMPP at the local level would be difficult.

4.2.3 Factor Analysis on Selected Set of Variables--

The remaining set of 13 variables was factored and analyzed and yielded the 5 factors shown in Table 4-8, which explain 65.7 percent of the variance in the variable set. Note, all of the variables now appear in the first 5 factors.

Factor 1 loads on all 5 central tendency measures and this is a measure of the proportion of activity which takes place in the central city (cities) of the SMSA. It is not surprising this appears as the primary indicator of urban form.

Factor 2 loads on 3 variables: portion of land which is urbanized (LU), portion of workers working outside the SMSA (TCOM), and proportion of land devoted to outdoor recreation (REC). It appears this factor may differentiate between SMSAs within and outside New England; the following cities have scores greater than 2.0 on this factor:

Bridgeport, Connecticut
Bristol, Connecticut
Brockton, Massachusetts
Danbury, Connecticut
Lowell, Massachusetts-New Hampshire
Meriden, Connecticut
Nashua, New Hampshire
New Britain, Connecticut
Norwalk, Connecticut
Stamford, Connecticut.

This is due in part to the fact that the recreation variable and the urbanized land variables were estimated from county data for New England.[a] In addition, loadings on LU and TCOM described an urban area which spreads beyond the boundaries of the SMSA, which is characteristic of New England's metropolitan areas.

---

[a] See documentation in Volume III.

Table 4-8

SCS2 - Stage I - Factor Analysis of 15 Variables

Variable	FACTOR 1	2	3	4	5
Portion of Employment in Central City	++				
Portion of Total Population in Central City	++				
Portion of Manufacturing Employment in the Central City	++				
Portion of Retail Sales in Central City	++				
Portion of Manufacturing in Central City	+				
Portion of Land Which is Urbanized		++			
Portion of Workers Working Outside SMSA		++			
Portion of Land Devoted to Outdoor Recreation		+			
Number of Radial Roadways			++		
Number of Major Circumferential Roadways			+		
Square Miles Per Person			-		
Total Miles of Roadway per Capita				++	
Portion of Principal Arterials				--	
Number of Population Centers					++
Square Miles per Person - Central City					+

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

This is due, in part, to the fact that the recreation variable and the urbanized land variables were estimated from county data for New England.* In addition, loadings on LU and TVUM describe an urban area which spreads beyond the boundaries of the SMSA, which is characteristic of New England's metropolitan areas.

Factor 3 has high positive loadings on the number of primary radial facilities (RAD) and the number of primary circumferential facilities (CIR) in the urban area. The factor loads negatively on square miles per person in the SMSA (SQMS). Together, these indicate a heavily urbanized area.

Factor 4 loads heavily on highway-miles per capita (HWPC) and negatively on proportion of principal highway facilities (HWPR). This is a simple factor which indicates the type of roadway network serving the area. The data utilized for the 2 highway variables is actually estimated by State officials rather than measured. Therefore, it is suspicious that all SMSAs with scores greater than 2.4 on this factor are located in Michigan. There are 9 such cities:

Ann Arbor, Michigan	2.5
Battle Creek, Michigan	2.8
Bay City, Michigan	2.5
Flint, Michigan	3.0
Jackson, Michigan	4.0
Kalamazoo-Portage, Michigan	5.4
Lansing-East Lansing, Michigan	4.7
Muskegon-Muskegon Heights, Michigan	5.8
Saginaw, Michigan	2.6

Obviously, the data were estimated and reported much differently for Michigan than for other States. It follows that the State to State differences in reporting may cause real distortions in the correlations over all observations and the factor scores for SMSAs. This factor will also be dropped.

Factor 5 loads heavily on the number of population centers in the area (PCR) and the number of square miles per person in the central city (cities) (SQMC). This indicates a dispersed pattern of urbanization involving more than one community in the core of the area.

These 5 factors together seem to describe the cities well; however, factors 2 and 4 are of questionable value since the correlations on which they are founded are largely the result of systematic data problems.

4.3 SCR3: Pollutant Residuals

Analysis of available data on pollution residuals was relatively simple due to the uniformity of the data base. Measures of the 12 pollutants listed in Table 3-5 were available for all but the following 13 SMSAs:

1170	Bristol, Connecticut
1200	Brockton, Massachusetts
1930	Danbury, Connecticut
2480	Fall River, Massachusetts-Rhode Island
2600	Fitchburg-Leominster, Massachusetts
4160	Lawrence-Haverhill, Massachusetts-New Hampshire
4560	Lowell, Massachusetts-New Hampshire
4960	Meriden, Connecticut
5350	Nashua, New Hampshire
5440	New Britain, Connecticut
5760	Norwalk, Connecticut
8040	Stamford, Connecticut
8760	Waterbury, Connecticut.

Note these are all in New England and the use of means will produce a bias in second stage factor analysis or clusters.

Following initial analysis of this data, 2 variables were eliminated from the set: noncombustible solid waste (RNSW) and industrial sludge (RIS). These 2 variables show very low correlations with each other and all water pollutants but seem to correlate well with air pollutants. The variables were eliminated because other SCAs contain no data related to land pollutants or their effects, therefore, the variables were not relevant.

Factors developed from the remaining data are shown in Table 4-9. The variables comprise one factor which measures air pollutants, one which measures water pollutants, and one which measures a combination of the 2. Together the 3 factors explain 67.5 percent of the variance in the 10 variables.

4.4 SCA4: Household Sector

Analysis of SCA4 data yielded factors which describe the socioeconomic character of an urban area. In this sense, the analysis of this SCA is more similar than any other to previous efforts in city classification. The analysis was begun with the variables shown in Table 3-6 plus 2 additional measures: latitude (LAT) and longitude (LONG). These 2 variables were added as temporary variables to help identify regional effects in the raw data and factors.

4.4.1 The Regional Effect--

Initial factor analysis produced 8 factors, as shown in Table 4-10. Several aspects of a metropolitan area are characterised by these factors: standard of living, housing, transportation, growth, employment, family structure, and health standard. This set of factors is dominated, however, by a regional effect in the first factor; note that 3 of 4 variables on which the factor loads most heavily are north-south indicators. The extent to which this effect dominates is obvious from the groups of SMSAs which form when cluster analysis is performed on these Stage 1 factors. Consider, for instance,

Table 4-9

SCB3 - Stage I - Final Factors

Variable	Factors 1	2	3
Sulfur Oxides	++		
Nitrogen Oxides	++		
Particulates	++		
Carbon Monoxide	+	+	
Wastewater		+	
Suspended Solids		+	
Hydrocarbons		+	
Nutrients			+
Biochemical Oxygen Demand			+
Dissolved Solids			+

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

Table 4-10

SC84 - Stage 1 - Initial Factors

Variable	Factors 1	2	3	4	5	6	7	8
Latitude	--							
Income -- Gini Coefficient	--							
Mean Temperature in January	++							
Mean Temperature in June	++							
Median Family Income				-				
Crowded Housing Units	+				+			
Married Couples without Own Household	+							
Per Capita Income	-			+				
Portion of Population Which is Black	+						+	
Units in Structure -- More than 3 units		++						
Single Family Dwellings		--						
Owner-Occupied Housing		--						
Units in Structure -- More than 50 Units		+						
Work Trip -- Drivers		+						
Work Trip -- Passengers		-						
Portion of Population which is of Foreign Stock		+						
Portion of Population which is Urban		+						
Households with 1 or more Automobiles		-	+					
Movers into the SMSA		++	++					
Population Change -- SMSA			++					
Individuals Residing in the Same Dwelling for 5 yrs.			--					
Female/Male in 20-64 Age Group			+					
Population Change -- Central City			+					
Median Age			-		-			
Relative Death Rate			-				+	
Employment/Population -- Total				++				
Employment/Population - Male				++				
Employment/Population -- Black				+				
Employment/Population -- Female				+				
Work Trip -- Walkers				-			+	
Average Household Size					++			
Fertility Rate					+			
Change in Median Income						++		
Longitude of SMSA						-		
Infant Death Rate							+	
Vehicle Miles Travelled Per Capita								+

Std. of living

Housing

Transportation

Growth

Employment

Family structure

health std.

78.9% of variance explained by these factors.

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

0040 Abilene, Texas
4200 San Angelo, Texas
9080 Wichita Falls, Texas
8600 Tuscaloosa, Alabama
8720 Fort Smith, Arkansas-Oklahoma

and

1360 Cedar Rapids, Iowa
6880 Rockford, Illinois
2760 Ft. Wayne, Indiana
8920 Waterloo-Cedar Falls, Iowa
3800 Kenosha, Wisconsin
6600 Racine, Wisconsin
3000 Grand Rapids, Michigan
7800 South Bend, Indiana
4320 Lima, Ohio
0440 Anderson, Indiana

It is clear from these examples that if geography is not to dominate this analysis, the data set must be cleared of all variables which are solely regional indicators.

4.4.2 Inactive Variables--

With this in mind, the next iteration of the analysis was performed without the following variables: latitude (LAT), longitude (LONG), mean temperature in January (TJA), and mean temperature in July (TJU). The resulting factors, shown in Table 4-11, describe the same characteristics as the previous iteration, however, the factors appear in different order and in a more consolidated form. This order of the factors is significant because it indicates the order of importance of the variables as determined by the amount of variance explained.

Note that the 3 variables which comprise factors 7 and 8 do not fall into the factor loading as would be expected. Percent of employed persons walking to work (TWK) and vehicle miles per capita (VMTP) separate themselves both from each other and from the factor which includes all other transportation variables. And, median income change (YMC) is isolated from employment, growth, and standard of living factors. The reason these variables do not "fall in" is apparently different for each. VMTP appears to be a poorly estimated variable, probably because the variable is derived from gasoline sales reports at the State level and local estimates of street-miles. In any case, the measure does not correlate well with any other measure in this SCS. TWK and YMC, on the other hand, correlate with a number of different measures in this SCS, to the extent that it appears TWK and YMC are determined by the combined influence of other independent factors in the set which are the primary dimensions of variance between SMSAs.

Table 4-11
ACR4 - Stage 1 - Second Iteration Factors

Variable	Factors 1	2	3	4	5	6	7	8
Units in Structure -- More than 5 units	++							
Single-Family Dwellings	--	Housing						
Owner-Occupied Housing	--							
Units in Structure -- More than 50 Units	+							
Portion of Population which is Urban	+							
Work Trip -- Passengers	+							
Portion of Population which is of Foreign Stock	+							
Work Trip -- Drivers	-	Transp (-TWK, VMTP						
Households with one or more Automobiles		+						
Movers into SMSA		++						
Individuals Residing in the Same Dwelling for 5 years		--	Growth					
Population Change -- SMSA		++						
Population Change -- Central City		+						
Female/Male in 20-34 Age Group		+						
Median Age		-			-			
Employment/Population -- Total			++			Family		
Employment/Population -- Male			++	Employment				
Employment/Population -- Female			+					
Employment/Population -- Black			+					
Income -- Gini Coefficient				++				
Median Family Income				--				
Per Capita Income		std. of living		-				
Crowded Housing Units				+	+			
Married Couples without Own Household				+				
Average Household Size					++			
Fertility Rate					+			
Portion of Population which is Black						++		
Infant Death Rate						+	health std.	
Relative Death Rate						+		
Work Trip -- Walkers							++	
Change in Median Income							+	
Vehicle Miles Traveled per Capita								+

79.2% of variance explained

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

4.4.3 City Size --

A diversion from the main course of analysis was entertained at this point to check one of the basic hypotheses by which this research has been performed. All data used in this research has been "normalized" to eliminate the dominating influence of city size. Most measures (in all SCSs) are expressed as percent of total or per capita rather than absolute values. The hypothesis is that the underlying effects of city size will not be eliminated by this procedure, but will be prevented from obscuring other major effects. A special factor analysis was performed to test this hypothesis.

The results (see Table 4-12) are for a factor analysis on the same variables as those shown in Table 4-11 and discussed in Section 4.4.2, except that total population of the SMSA (P) is included. The effect of the addition of this variable is not dramatic: growth, housing, employment, standard of living, transportation, family structure, and standard of health still appear in the factor set, although in slightly different form. The major change is in the factor which describes location and density. In the first instance, housing and transportation factors combine to form one such factor. When the population variable is added, however, it appends itself to the housing factor and replaces transportation measures which are dispersed throughout the factors. As the transportaton measures utilized are largely a function of city size and housing patterns, it seems the location-density factor and the factor structure are not really changed with the new variable combinations.

The fact that P does not form a primary factor of its own indicates the effect of city size has already been included in the data set. And, because P associates strongly with one and only one of the previous factors, it is clear the effect of city size has been isolated to some extent rather than being allowed to infiltrate throughout the factor set.

4.4.4 Final Stage I Factors --

The final iteration in Stage I for this SCS simply eliminated those variables identified as inactive in Section 4.4.2 (TWK, YMC, and VMTT) since they were not serving a useful function in the factor description of SMSAs. Surprisingly, the elimination of these variables caused the formation of a new factor which absorbed several tag-along variables from other factors to comprise an indicator of cosmopolitan character. The results of the final iteration are shown in Table 4-13 and summarized by Table 4-14.

4.5 SCS5: Public Sector

As indicated in Table 3-7, there are 10 variables in the data set to describe the public sector. Sunday newspaper circulation per capita (NDW) and presidential votes cast per thousand population (VOT) are the only measures not taken from the Census of Government.

Table 4-12
SCSA - Stage I - Test of the Effect of City Size

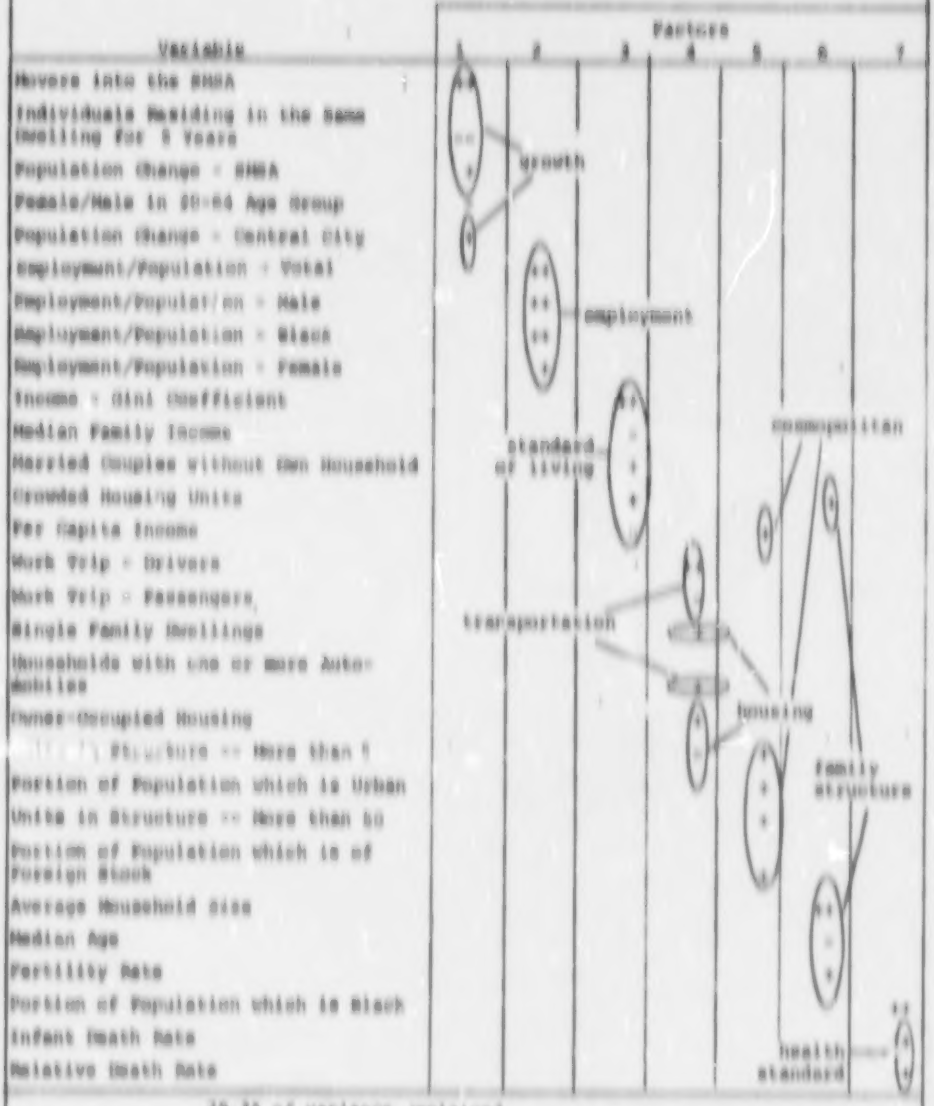
Variable	Factor 1	2	3	4	5	6	7	8
Movers into the SMSA	++							
Individuals Residing in the Same Dwelling for 5 years	--							
Population Change -- SMSA	++							
Population Change -- Central City	+	Growth						
Female/Male in 20-64 Age Group	-							
Median Age	-					-		
Households with one or more Automobiles	+							
Units in Structure -- more than 50 units		++						
Population		++						
Portion of Population which is Urban		+	Housing					
Units in Structure -- more than 5 units		+						
Owner-Occupied Housing		-						
Single Family Dwellings		-			-			
Work Trip -- Passengers		+						
Employment/Population -- Total			++					
Employment/Population -- Female			++	Employment				
Employment/Population -- Male			+					
Employment/Population -- Black			+				Family	
Income -- Gini Coefficient				++				
Median Family Income				--				
Crowded Housing Units		standard of living		+		+		
Per Capita Income				-				
Married Couples without Own Household				+				
Work Trip -- Walkers					++			
Work Trip -- Drivers					--			
Average Household Size						++		
Fertility Rate				transportation		+		
Portion of Population which is Black							++	
Infant Death Rate							+	health std.
Relative Death Rate							+	
Vehicle Miles Traveled per Capita								+
Portion of Population which is of Foreign Stock								
Change in Median Income								

78.6% of variance explained.

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

Table 4-13

SCS4 - Stage I - Factors


growth
employment
standard of living
cosmopolitan
transportation
housing
family structure
health standard

Variable	Factors 1	2	3	4	5	6	7
Movers into the SMSA	++						
Individuals Residing in the Same Dwelling for 5 Years	--						
Population Change - SMSA	+						
Female/Male in 20-64 Age Group							
Population Change - Central City	+						
Employment/Population - Total		++					
Employment/Population - Male		++					
Employment/Population - Black		++					
Employment/Population - Female		+					
Income - Gini Coefficient			++				
Median Family Income			-				
Married Couples without Own Household			+				
Crowded Housing Units			+			+	
Per Capita Income			-		+		
Work Trip - Drivers				+			
Work Trip - Passengers				-			
Single Family Dwellings				+			
Households with One or More Automobiles				+			
Owner-Occupied Housing				+			
Units in Structure -- More than 1				-	+		
Portion of Population which is Urban					+		
Units in Structure -- More than 50					+		
Portion of Population which is of Foreign Stock					+		
Average Household Size						++	
Median Age						-	
Fertility Rate						+	
Portion of Population which is Black							++
Infant Death Rate							+
Relative Death Rate							+

78.3% of variance explained

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

Table 4-14

SCB4 - Stage I : Interpretation of Final Factors

Factor	Positive Score Indicates some Combination of
1. Growth	Immigration to the SMSA Increasing population in SMSA and central city
2. Employment	High employment participation rates in all sectors of the population
3. Standard of Living	Relatively low level of income as indicated by median and per capita measures Unequal distribution of income over the population Crowded housing
4. Location-Density	High proportion of single unit and owner-occupied housing Heavy dependence on the automobile
5. Cosmopolitan	Relatively high level of income as indicated by a per capita measure Dense housing Urban and foreign elements of the population relatively large
6. Family Structure	Crowded housing Large proportion of young families
7. Health Standard	High infant death rate Relatively high overall death rate given the age distribution of the population.

4.5.1 Preliminary Analysis --

Table 4-15 shows the factors resulting from preliminary factor analysis on these variables. Although these factors are not as meaningful as those developed for other SCEs, since most of the variables reflect levels of expenditure, distinct aspects of the local government can be identified. Factor 1 loads heavily on 3 variables which describe the overall level of government operations relative to city size: total revenue per capita (GREV), total expenditures per capita (GX), and employment per capita (GEMP). Factor 2 includes 2 variables measuring sewerage expenditures. Factor 3 indicates expenditures on water supply and expenditures on parks and recreation are highly correlated. And, Factor 4 identifies high expenditures on solid waste collection and low voting participation.

Although both NEW and VOT appear in factors with eigenvalues greater than one, it appears they really do not belong in this factor set. VOT is most highly correlated with GREV (.26) and NEW correlated most highly with VOT (.12)--but neither correlated highly with any variable in the set. NEW is included in Factor 2, which indicates the general level of expenditures on sewerage, but does not intuitively "fit." VOT is one of 2 variables which comprise factor 4 but the combination of VOT with GXSO (expenditures for sanitation other than sewerage) makes sense only to the extent that public-spirited people may not litter. In addition, there is some question as to the relevance of the variable NEW, since it represents Sunday newspapers rather than daily newspapers and the total number published in an SMSA rather than the total number purchased and read. As these variables did not make a substantial contribution either hypothetically or statistically, they were eliminated from the variable set before the final factor analysis.

4.5.2 Final Factors --

The elimination of NEW and VOT from the data set did not change the resulting factor structure. As Table 4-16 shows, the only change in loadings is the deletion of the rows corresponding to those two variables.

Table 4-15

SCBS - Stage 1 - First Iteration

Variable	Factors 1	2	3	4
Local Government				
General Revenues	++			
General Expenditures	++	large government		
Employment	++			
Sewerage Expenditures -- Total		++		
Sewerage Expenditures -- Other than Capital Outlay		+	sewerage	
Sunday Newspaper Circulation		+		
Expenditures on Water Supply			++	
Expenditures on Parks & Recreation			+	water & parks
Expenditures on Sanitation other than Sewerage				++
Presidential Votes Case per Voting Age Population				-

69.9% of variance explained

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

Table 4-16
SCS5 - Stage I - Final Factors

Variable	Factors 1	2	3	4
Local Government				
General Revenues	++			
General Expenditures	++ — large government			
Employment	++			
Sewerage Expenditures -- Total		++		
Sewerage Expenditures -- other than capital		++ — sewerage		
Expenditures on Water Supply			++ — water & parks	
Expenditures on Parks & Recreation			+	
Expenditures on Sanitation other than Sewerage				sanitation — ++

85.5% of variance explained.

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--). The factors are numbered in order according to the percent of variance explained.

4.6 SCSS: Industrial Sector

As this analysis is part of a classification which is oriented toward environmental quality, initial data collection and analysis of industry was oriented in that direction as well. The Bureau of the Census provided data which reflected local levels of production.

4.6.1 Analysis of SEAS Data --

SEAS data included total value of production in approximately 185 INFORUM sectors* and was readily available from the SEAS data bank. Sectors were selected from this set based on pollution generation and abatement coefficients in order to obtain data for those industries with especially high or low pollution generation or especially high or low abatement. Total value of production in each of these sectors plus value added in all manufacturing and total employment in manufacturing provided 35 variables for the initial factor analysis. Variables which grouped together to form the first 8 factors are shown in Table 4-17. (A total of 15 factors which formed had eigenvalues greater than one.) Since factors identify sectors which locate together in different urban areas and the set of industries represented is based on pollution and abatement coefficients, the factors are more statistically than conceptually meaningful. Even statistically, however, they are based on very weak correlations.

Several other problems detract from the usefulness of these factors. First, levels of production are described in terms of value of product rather than value added; since the ratio of these 2 measures can vary substantially within a sector, value added is a more accurate measure of the actual level of production. And, second, industrial activity as reported by SEAS is not actual levels for SMSAs. Instead, this data is estimated by sharing out forecast, national activity levels by sector between SMSAs according to a forecasted distribution of industrial activity.

Given the above caveats for utilizing SEAS industry SMSA data, additional data was collected from the Census of Manufacturing. This data included value added by each of the 19 two-digit SIC sectors--actual levels in 1972 by SMSA. The next step, then, was to compare the two bodies of data. Since the two sources used different sector categories, this comparison is performed indirectly through factor analysis. Table 4-18 lists the industry sectors from SEAS and the Census which group together and confirms the validity of SEAS data. Since the Census data is no less descriptive, and now up on the computer, this body was used for the remaining analyses.

4.6.2 Industry Mix as Determined from Census Data--

Input to the final factor analysis included total value added in manufacturing and total employment in manufacturing as well as value added in each of 19 sectors of the economy. As indicated above, these sectors are defined by the two-digit SIC code and represent the overall mix of industry rather than relying on hypotheses and estimates of pollution or abatement by individual sectors.

---

*See Volume III for a complete discussion of this data and its use.

Table 4-17

SCS6 - Stage I - Initial Factor Analysis

Factor	Sector	
1	IPRT	Commercial Printing
	IWST	Wholesale Trade
	IPNT	Paints
2	IELC	Electric Utilities
	ISTL	Steel
	ICOL	Coal Mining
	ICLY	Structural Clay Projects
	IMTL	Fabricated Metal Products
3	EM	Employment in Manufacturing
	ICAR	Motor Vehicles
	VAM	Value added in Manufacturing
	IOIL	Crude Petroleum, Natural Gas (-)*
4	IPLA	Plastic Materials and Resin
	IFBR	Cellulose Fibers
5	IFUL	Petroleum Refining
	ICHM	Industrial Chemicals
6	IGLS	Glass
	IVEG	Canned and Frozen Foods
7	IPB	Lead
	ICU	Copper
8	IABS	Stone and Clay Products
	IFRN	Household Furniture

*) 15 factors explain 64%.

Table 4-18

NCSS - Comparison of SEAS and Census Industry Data

Factor		Sector	
1	Census	S34	Fabricated Metal Products
	Census	S35	Machinery, except electrical
	SEAS	IMTL	Other Fabricated Metal Products
2	Census	WHOL	Wholesale Sales
	SEAS	IWST	Wholesale Trade
	Census	S27	Printing and Publishing
	SEAS	IPRT	Commercial Printing
3	SEAS	ICHM	Industrial Chemicals
	Census	S28	Chemicals and Allied Products
	Census	S29	Petroleum and Coal Products
	SEAS	IFUL	Petroleum Refining
4	SEAS	IGLS	Glass
	Census	S32	Stone, Clay and Glass Products
	SEAS	IVEG	Canned and Frozen Foods
5	SEAS	ICTN	Broad and Narrow Fabrics
	Census	S22	Textile Mill Products
6	SEAS	ICOL	Coal Mining
	SEAS	IELC	Electric Utilities
	SEAS	ICLY	Structural Clay Products
7	SEAS	IFRN	Household Furniture
	Census	S25	Furniture and Fixtures
8	Census	S26	Paper and Allied Products
	SEAS	IPPR	Paper and Paperboard Mills
	SEAS	IPLA	Plastics Materials and Resins

Results of the final factor analysis are shown in Table 4-19. The 8 factors described here explain 59 percent of the variance indicated by the 24 input variables. These 8 factors appear to describe well the basic dimensions of industry mixes and could be easily titled as follows:

- overall level of industrial activity
- services and trade
- textiles and apparel
- miscellaneous manufacturing and instruments
- fuel and chemicals
- paper and allied products
- leather and leather products
- lumber and wood products.

The factors are listed in order according to the amount of variance explained by each; the latter factors, therefore, are the least valuable to the factor description of SMSAs.

Table 4-19

SCGS - Stage I - Final Factors

Factor		Variables
1	VAM	Value added in Manufacturing
	EM	Employment in Manufacturing
	S34	Value added in Fabricated metal products
	S35	Value added in machinery, except electrical
2	WHOL	Wholesale Sales
	RETT	Retail Sales
	SS	Selected Services
	S27	Value added in printing and publishing
3	S22	Value added in textile mill products
	S23	Value added in apparel and other textile products
4	S39	Value added in miscellaneous manufacturing industries
	S38	Value added in instruments and related products
5	S28	Value added in chemicals and allied products
	S29	Value added in petroleum and coal products
6	S26	Value added in paper and allied products
7	S31	Value added in leather and leather products
8	S24	Value added in lumber and wood products

5.0 STAGE II FACTOR ANALYSIS

The two-stage application of the factor analytic technique was utilized to insure the proper evaluation of research hypotheses. Stage I analysis involved a separate factor analysis for each SCS to reveal the hypothesized underlying dimensions within each group of variables--each of these is described in detail in Chapter 4. The output from individual Stage I analyses was then combined and used as the input to Stage II. Stage II thus identifies the relationships between SCSs and the basic underlying attributes of U.S. metropolitan areas.

This chapter narrates the analytic process and results of the Stage II analysis. Section 5.1 begins with an evaluation of alternative methods of rotation for the Stage II factors. Section 5.2 then describes the selection of Stage I factors to be included in Stage II analysis. Results of Stage II are presented and interpreted in Section 5.3. And the stability of factors between city sizes is investigated in Section 5.4. The use of Stage II factors to classify cities is documented in the following chapter.

5.1 Alternative Methods of Factor Rotation

Stage I factors have been constrained to be independent of each other (orthogonal); the reason for this was to emphasize the relationships between, as opposed to within, SCSs in Stage II analysis. This constraint does not decrease the amount of variance explained by the factors but may make them slightly more difficult to interpret. The desired product of Stage II, however, is the set of factors which most clearly describes U.S. metropolitan areas--regardless of the interrelationships of the factors. The first step in Stage II, then, was to investigate the advantages of an oblique rotation over an orthogonal rotation.

Figure 5-1 illustrates the potential difference between the methods using a two-dimensional example. The points A, B, C, D, and E on the diagrams represent variables, which are located on the factor axes by factor loadings. Note that the orthogonally rotated factors have much simpler factor loadings than the unrotated factors and are, therefore, more easily interpreted.* Since the oblique rotation is not limited to independent factors, however, the oblique factors come still closer to the variables, factor loadings are simplified slightly more, and the factors are still easier to interpret.

Preliminary application of the factor analysis alternatives to Stage II indicated the difference would be much less dramatic--in fact, not significant. Twelve Stage I factors were combined for Stage II analysis and the resulting Stage II factors rotated according to three methods: orthogonal, direct oblique, and indirect oblique. The difference in factor loadings between the three methods is so small that all three matched the pattern shown in Table 5-1; and the proportion of variance explained in each case was 60.5 percent.

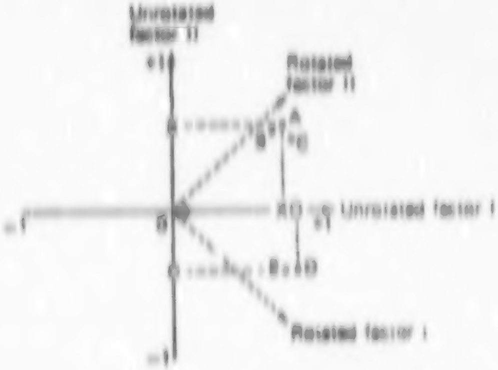
---

*e.g.: The unrotated Factor 1 loads moderately on all variables whereas the rotated factor 1 loads heavily on D and E and has only small loadings on A, B, and C.

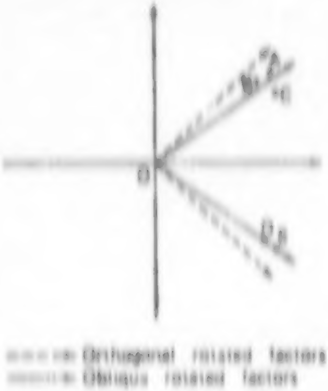
Figure 5-1

Alternative Methods of Factor Rotation

Orthogonal Factor Rotation


Unrotated factor II
+1
Rotated factor II
A
B
+C
-1
0
Unrotated factor I
+1
D
Rotated factor I
-1

Oblique Factor Rotation


A
B
+C
0
D
Orthogonal rotated factors
Oblique rotated factors

SOURCE: Nie, Bent & Hull, Statistical Package for the Social Sciences, New York: McGraw Hill Book Company, 1970, p. 222 (Figure 3) and p. 223 (Figure 4).

Table 5-1

Preliminary Stage II Factors*

SCS	Stage I	Factor	Stage II Factor 1	2	3	4	5
5	2	Expenditure for Sewerage	+				
4	3	Standard of Living (low)	-				
2	2	Highly Urbanized	+				
5	4	Expenditure for Sanitation other than Sewerage		++			
2	3	Street-miles per capita, proportion of local streets		-			
4	4	Location-density		-			
5	3	Expenditure-Water Supply, Recreation			+		
2	4	Sprawling Core			+		
2	1	Central Tendency				++	
4	1	Growth				+	
5	1	Large Government					-
4	2	Employment					+

*Note: SCS 6 is not represented here because the Stage I factors were not available at this point. The absence of SCS1 and SCS3 is explained in Section 5.2.

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); -.75 to -1.00 (--).

As the oblique rotation of factors did not promise a significant advantage in interpretation and yet involves greater expense on computer analysis, all further rotations were constrained to orthogonal factors.

5.2 Selection of State I Factors to be Included in Stage II

For a variety of reasons--both conceptual and statistical, a select set of Stage I factors was chosen for input to Stage II. This set included the first 4 factors from 4 SCSs--2, 4, 5, and 6--with the exception of factor 3 from SCS2. The process of elimination is described below.

5.2.1 Environmental Quality and Causation--

SMSAs are to be classified in groups which would react similarly to environmental policies. The 6 significant characteristic sets described above include both environmental quality and those attributes which affect environmental quality--environmental quality and residuals are largely endogenous to other urban systems. Thus, two alternatives present themselves, with respect to grouping cities for this research. The first is to use the data base composed of all 6 SCSs, including environmental quality and residuals. The resulting groups would reflect both basic metropolitan area characteristics and present environmental quality. The second alternative is to classify cities on basic attributes of the urban area--excluding environmental quality and residuals. Since environmental quality is largely a function of these other basic characteristics of the city, the groups yielded by the two alternatives should be similar.

SCS1 and SCS3 contain all measures of environmental quality. In the course of working with this data it has been determined that the data in these two sets is substantially weaker than the rest of the data base in several respects. This has been discussed in Chapters 3 and 4.

Considering the reduced set of cities for which data is available and the other data problems faced, the decision was made to perform the factor analysis and clusters on the 4 data sets which represent characteristics of the city exogenous to environmental quality. Limiting the data base to nonenvironmental measures of urban character has several advantages. The number of cities which can be classified is expanded. Bias inherent in existing data on environmental quality is avoided. And, the classification is more versatile since it is derived from comprehensive measures of the demography, economy, government, and physical form of the community; the groups might thus be used for urban studies other than those related to environmental quality.

Chapter 6 describes the reintegration of environmental variables--groups of cities generated from Stage II factors are tested for their ability to discriminate between levels of environmental quality.

5.2.2 Balance Between the SCGs --

Stage I analysis yielded a different number of factors for each of the four SCGs to be pursued in Stage II. For SCGs 2, 4, 5, and 6, the number of factors with eigenvalues greater than one was 4, 7, 4, and 8, respectively, as indicated in Chapter 4. To use this set of factors in Stage II would create a significant bias toward SCG4 (the household sector) and SCG6 (the industry sector). In addition, since Stage I factors from each SCG are mutually independent, excess factors in SCG4 and SCG6 will lead to the formation of additional separate factors beyond the primary factors indicated by interactions between the 4 SCGs. With these considerations, the number of factors from Stage I to be included in Stage II was limited to 4 per SCG.

The loss of explanatory power from the exclusion of these factors is significant but tolerable. Table 5-2 shows the proportion of variance explained by each factor and indicates the loss will be 28.3 percent in SCG4 and 21.0 percent in SCG6. Percent of variance explained by the first 4 factors if each SCG is as follows:

2	Urban Form	73.4 percent
4	Household Sector	50.0 percent
5	Public Sector	81.5 percent
6	Industry Sector	38.0 percent

Although several alternative combinations of Stage I factors were tested, the even distribution of factors between SCGs yielded the most meaningful factors, therefore, the reduced set was accepted as the best input to Stage II.

5.2.3 Elimination of Undesirable Factors --

Two of the Stage I factors identified from urban form data (SCG2) are unique in that they characterize particular areas of the country rather than the basic attributes of urban areas. These two factors were described in Chapter 4. The first is factor 2, which seems to identify SMSAs in New England; the second is factor 4 which has very high loadings on Michigan's SMSAs exclusively. To some extent, these occurrences are explained by peculiarities of the data set (see Section 4.2). In any case, these two factors are not useful in explaining the underlying characteristics of urban areas and are excluded from further analysis, as stated in Chapter 4.

5.2.4 Input to Stage II Analysis --

At this point, having described the output from each SCG of the Stage I analysis separately and having eliminated several Stage I factors from the analysis, it may be helpful to review the factors to be used as input to Stage II. Table 5-3 identifies the factors selected for Stage II and provides a concise description of each to aid in the interpretation of Stage II factors.

5.3 Results of Stage II Factor Analysis

Factor analysis of the 14 Stage I factors shown in Table 5-3 yielded 6 Stage II factors which, together, explain 63.1 percent of the variance in Stage I factors. These factors are intuitively satisfying as well as statistically valid in that each factor represents a set of characteristics which are likely to be encountered together in an urban area.

Table 5-2

Proportion of Variance Explained by Stage 1 Factors (%)

Factor	SCE2*	SCE4	SCE5	SCE6
1	35	16	37	14
2	14	12	20	9
3	12	11	15	8
4	9	11	14	7
5		10		7
6		10		5
7		8		5
8				5
All Factors	70.4	78.3	85.5	59.0
First 4 Factors	70.4	50.0	85.5	38.0
Difference		28.3		21.0

*Note there are only 4 factors shown for SCE2 as opposed to the 5 documented in Chapter 4. The second of those described in Chapter 4 has been eliminated at this point since the factor was primarily differentiating between SMSAs inside and outside New England.

Table 5-1

Input to Stage II Factor Analysis (Stage I Factors)

Table	SCs	Factor	Title
4-8	2	1	Central tendency
		2	Highly urbanized
		4	Sprawling core
4-13	4	1	Growth
		2	Employment
		3	Standard of living (low)
		4	Location-Density (single family homes + high auto dependence)
4-16	5	1	Large government
		2	Expenditure on sewerage
		3	Expenditure on water supply and recreation
		4	Expenditure on sanitation other than sewerage
4-20	6	1	Overall level of industrial activity
		2	Services and trade
		3	Textiles and apparel
		4	Miscellaneous manufacturing and instruments

Table 5-4

Stage II Factors

PCS	Stage 1 Factor		Stage II Factor 1	2	3	4	5	6
4	3	Low Standard of Living	++					
5	2	Expenditure on Sewerage	-					
6	1	Overall Level of Manufacturing	-	-				
4	5	Growth		++				
2	1	Central Tendency		+				
6	4	Miscellaneous Manufacturing & Instruments			++			
2	4	Sprawling Core			-			
5	6	Expenditure on Non-Sewerage Sanitation			+			
4	2	Employment				++		
6	2	Services & Trade				++		
4	6	Location-Density (many single-family homes, high auto dependence)					+	
6	3	Textiles & Apparel					-	
5	3	Expenditure on Water Supply & Recreation					+	
5	1	Large Government						++
2	3	Highly Urbanized						-

Factor loadings are indicated as follows: .50 to .75 (+); .75 to 1.00 (++); -.50 to -.75 (-); and -.75 to -1.00 (--).

The factors are best described by the variables which comprise them; Table 5-4 depicts the composition of the factors by their loadings. As anticipated, Stage II factors identify correlations between the attributes identified in the separate SCSs with each Stage II factor loading heavily on one factor from each of 2 or 3 SCSs. Although each SCS is represented in at least 3 and not more than 5 factors, it appears industrial activity is represented most strongly and urban form has the least emphasis. This is not surprising since the data in SCS2 does not represent the research hypotheses well and the variables in SCS6 are extremely good indicators of city types (it seems reasonable that the character of a city should be closely related to the mix of industrial activity in the area).

Factor 1 indicates a low income, low government expenditures for sewerage, and a low level of total manufacturing activity; this factor would characterize an economically depressed area. Extreme cities on this factor scale were*

+ McAllen-Pharr-Edinburg, Texas
+ Brownsville-Harlingen-San Benito, Texas
- Ann Arbor, Michigan
- Flint, Michigan

Both McAllen and Brownsville are border towns and as such, host a large number of immigrants living under relatively poor conditions. It is interesting to note that this factor appeared as the first factor in separate analyses for small, medium and large cities--the factor is extremely stable for all city size categories (see Section 5.4 below).

Factor 2 indicates a low level of total manufacturing activity, rapid growth in recent years and a high concentration of population and economic activity in the core (central city) of the SMSA. Extreme observations on factor 2 were

+ Austin, Texas
+ Lawton, Oklahoma
- Youngston-Warren, Ohio
- McAllen-Pharr-Edinburg, Texas

Austin, Texas, is the home of Bergstrom AFB, the Texas State capitol, and the main campus of the University of Texas. Together these institutions employ almost 42 percent of the total labor force (52,800). Lawton, Oklahoma, is the home of Fort Sill (Army) which, by itself employs almost 52 percent of the labor force (261,000). The factor seems to indicate a large SMSA with a nonindustrial economic base experiencing rapid growth.

---

*NOTE: + = high positive factor score
- = high negative score.

Factor 3 indicates manufacturing activity in the miscellaneous category, a compact core and heavy expenditures for sanitation other than sewerage. Extreme cities on the factor were

+ Providence-Warwick-Pawtucket, Rhode Island
+ Waterbury, Connecticut
− Battle Creek, Michigan
− Muskegon-Muskegon Heights, Michigan.

The largest industries in Providence are jewelry and silverware, and Waterbury specializes in brass products, plastics, and electronics. The factor seems to identify cities which are active in light industries.

Factor 4 indicates high employment and service and trade activities. Extreme cities on this factor were

+ Atlanta, Georgia
+ Anaheim-Santa Ana-Garden Grove, California
− Fayetteville, North Carolina
− Lawton, Oklahoma

Atlanta is the most illustrative of this group since it is the financial, retail, and wholesale center of the Southeast. It is the transportation hub, as well, served by six rail systems and the second busiest airport in the U.S.

Factor 5 indicates low residential densities, high auto dependence, little manufacturing in industries such as textiles and apparel, and heavy expenditures on water supply and recreation. Extreme observations on this factor were

+ Las Vegas, Nevada
+ Anaheim-Santa Ana-Garden Grove, California
− Jersey City, New Jersey
− Fall River, Massachusetts

Las Vegas and Anaheim demonstrate the spreading land use and lack of industry denoted by this factor. Jersey City and Fall River, on the other hand, are dominated by manufacturing--textiles in particular. All evidence indicates that this factor is describing suburban attributes.

Finally, factor 6 indicates a highly urbanized SMSA with a relatively large local government. Extreme observations on this factor were

+ New York, New York
+ Stockton, California
− Meriden, Connecticut
− Waterbury, Connecticut

Urbanized, in this case, means high population density and a strong radial pattern in roadways. This is the less important of the two stage I factors involved. Large local government means high per capita levels of revenue, expenditures, and employment by the local government. New York, of course, has an extremely large government since it serves a much larger number of people than those who live in the SMSA. Stockton may be in a similar position on a much smaller scale since it is a regional retail business center and is involved in the regional water authority which supplies the entire area, including San Francisco. Meriden and Waterbury are both specialized manufacturing communities and it is understandable that property taxes, expenditures on education, etcetera might be lower than normal in such communities. Conceptually, this last factor is much weaker than all others.

These stage II factors appear to be most satisfactory in every sense (with the exception of factor 6) with respect to the research hypotheses and ex poste examination of the variable combinations. The application of these factors to cluster analysis will provide still another test of their power in discriminating between SMSAs of inherently different character.

5.4 Analysis of Small, Medium and Large Cities

Before proceeding to Stage III of the analysis, which identified groups of similar cities, a test was performed to determine the stability of Stage II factors between different size cities. The factor analysis was repeated for each of three groups of cities:

Group	Population
small	less than 250,000
medium	250,000-500,000
large	more than 500,000

The proportion of variance explained by primary factors is stable only varying from 61.7 percent to 65.4 percent.

The differences between all-city factors and factors from these strata of cities are worth noting, however. Factor 1 from the all-city analysis (see Table 5-4) shows up as factor 1 in the analysis of each group of cities. Factors 2 and 3 (all cities) reappear only in the medium size city analysis--factor 2 is the second in explanatory power in this category as well, but factor 3 ranks last in importance. Factor 4 appears in each analysis, but appears to be very important for small cities: this factor ranks second, third, and fourth in the small, medium and large city analyses, respectively. Finally, factor 5 appears fourth in the medium city analysis and second in the large city analysis. The fact that these factors shift positions between the different strata of cities but do appear repeatedly lends further support to the factors as a set of measures capturing the underlying attributes for all sizes of cities. The shift in importance of the factors is an indication of the inherent differences between cities of different size.

6.0 SMSA GROUPS BASED ON GENERAL CHARACTERISTICS

The initial application of this research was to identify city groups and modal cities suitable for general environmental analysis. Five major city groups were found showing similarities in terms of the characteristics described by Stage II factors.

The classification represents information from SCS2 (urban form), SCS4 (households), SCS5 (public sector), and SCS6 (industry). As explained in Chapter 5, there are two reasons for this: first, according to the conceptual design, residuals (SCS3) and environmental quality (SCS1) are both results of the attributes and activities described by the other sets; and, second, environmental and residuals data were less reliable than the rest of the data bank. Analysis of the groups described here using environmental indicators is the subject of Chapter 7.

6.1 The Classification Process

Factor scores from Stage II analysis describe the 262 SMSAs in a multidimensional space. It is from the relative positions of these observations (SMSAs) in the space that SMSA groups and modal cities are identified.

The grouping algorithm utilized is a simple geometrical clustering technique:* each observation (SMSA) is initially considered a separate group. The two groups separated by the smallest geometric distance are then located and combined to form a new group with its centroid midway between the two points. Then, the two groups with nearest centroids of the new set are combined; and a new centroid located; the process continues until only one group remains.

Meaningful groups of cities are identified by retracing the grouping process and observing the point at which an optimal level of grouping has occurred. Criteria for choosing a stopping point include the size and number of groups and the relationship between within-group variance and between-group variance.

Once the preferred set of groups is selected, modal cities are identified by simply determining which city in each group lies closest in the multidimensional space to the geometric centroid of the group.

6.2 Input to the Grouping Algorithm

Theoretically, it would have been possible to develop SMSA groups directly from the initial variables using the geometrical clustering routine. In practical terms, however, the variables of the data base provide too many dimensions along which cities may differ--the additional descriptive information provided by the variables stresses the uniqueness of each city rather than underlying basic characteristics which the cities have in common. Because the use of too many dimensions creates an unmanageable set of groups, input to the cluster analysis was arbitrarily limited to 4 Stage II factors. The explanatory power lost at this point is significant: the portion of variance in Stage I factors explained by Stage II factors is 45 percent with 4 Stage II factors, as opposed to 63 percent with 6 Stage II factors. The effect on city groups of having excluded these last 2 factors is discussed in Section 6.4 below.

---

*Computer program written by Howard Gilbert and Steve Chasen, Health Sciences Computing Facility, University of California, Los Angeles, California. Reference: R.R. Sokal and P.H.A. Sneath (1973) Numerical Taxonomy: the Principles and Practice of Numerical Classification (San Francisco: W.H. Freeman and Co.).

6.3 SMSA Groups and Modal Cities

Table 6-1 shows how the 262 SMSAs grouped together to form basic classes of SMSAs. Five major groups of cities were identified; these include 175 of the 262 cities. The remaining cities grouped together as follows:

36	SMSAs in groups of 5 or more
34	SMSAs in groups of 2 to 4
17	SMSAs in groups of 1

These groups capture 81.2 percent of the total variance in factor scores between cities. The groups are not perfectly balanced in size, but do appear to differentiate between basic city types very well. A quick scan of the list confirms this.

In Table 6-1, the double lines indicate divisions between groups and the dotted lines delineate subgroups. To see the similarities and differences between groups and subgroups, refer to Table 6-2. The first 2 blocks of cities are separate subgroups from group 1; these cities are very similar except that factor 2 scores increase somewhat and factor 3 scores become more negative between subgroup 5 and subgroup 6. The sample subgroup from group 4, however, is very different. Between group 1 and group 4, the position on every axis changes dramatically. These are only examples, but are indicative of the discriminating power of the clustering algorithm.

Interpretation of the factor scores yields a clear impression of why these groups formed and what they mean. Table 6-3 characterizes each of the groups by identifying high scores on Stage II factors.

Group 1 shows a low level of manufacturing activity, low standard of living, and low expenditures on sewerage. This is tempered by moderate loadings on Stage II factors 3 and 4 which signify a growing economy oriented more than the average toward services and trade. Little Rock, the modal city for this group, is very close to the group's centroid (see Table 6-4). In 1970, Little Rock was less wealthy than the average SMSA as is indicated by the number of families below the low income level: 13.5 percent as compared to the national level of 8.5 percent.[*] But the area is growing and, in 1970, enjoyed a high

---

[*] This low income level was designated by the Social Security Administration in 1964 and modified by a Federal Interagency Committee in 1969. See p. XXXVIII of the 1972 County and City Data Book for a brief discussion of its derivation and meaning.

Table 4-1

Groups of Similar SMSAs*

GROUP 1

Abilene, TX
Lafayette, LA
San Angelo, TX
Midland, TX
Lubbock, TX
Tacoma, WA
Albany, GA

Brownsville, TX
West Palm Beach-Boca Raton, FL
Tampa, FL
Monroe, LA
Orlando, FL

Sherman-Denison, TX

Corpus Christi, TX
Shreveport, LA
Macon, GA
Texarkana, TX-AR
Wilmington, NC
Savannah, GA
Tyler, TX
Montgomery, AL
Jackson, MS

New Orleans, LA
Portland, ME
Waco, TX
Little Rock-N. Little Rock, AR**
Odessa, TX
Tulsa, OK
Baton Rouge, LA

St. Joseph, MO
Sioux City, IA-NE
Billings, MT
Boise City, ID
Huntsville, AL
Springfield, MO
Danville, VA

GROUP 2

Huntington-Ashland, WV-KY-OH
Stockton, CA
Modesto, CA

Augusta, GA-SC
Pueblo, CO
Fresno, CA

Killeen-Temple, TX
Lewiston-Auburn, ME
Pine Bluff, AR
Spokane, WA
Owensboro, KY
Fort Myers, FL
Yakima, WA

Florence, AL
Santa Rosa, CA
Pensacola, FL
Riverside-San Bernardino-Ontario, CA
Provo-Orem, UT
Charleston, SC
Salem, OR
Duluth-Superior, MN-WI

Altoona, PA
Gadsden, AL
Lake Charles, LA**
Mobile, AL
Bakersfield, CA
Chattanooga, TN-GA
Lakeland-Winter Haven, FL
Birmingham, AL

Erie, PA
Wheeling, WV-OH
Poughkeepsie, NY

Louisville, KY-IN
Richland-Kennewick, WA
Springfield, OH
Parkersburg-Marietta, WV-OH
Sacramento, CA

Gastonia, NC
Salt Lake City, UT
Petersburg-Colonial Heights-Hopewell, VA
Charleston, WV
Wilkes-Barre-Hazleton, PA
Beaumont, TX
Spartanburg, SC
New Bedford, MA
Alexandria, LA

MIXED GROUP

El Paso, TX
Tuscaloosa, AL
San Antonio, TX
Columbus, GA-AL

MIXED GROUP

Galveston-Texas City, TX
Manchester, NH
Santa Barbara, CA
Columbia, SC

GROUP 3

Allentown-Bethlehem-Easton, PA-NJ
Harrisburg, PA
St. Louis, MO-IL
Greenville, SC
Reading, PA
York, PA
Lancaster, PA

Cedar Rapids, IA
Waterloo-Cedar Falls, IA
Fort Wayne, IN
Toledo, OH-MI
Rockford, IL
Grand Rapids, MI

Davenport-Moline-Rock Island, IA-IL
Williamsport, PA**
Elmira, NY
Mansfield, OH
Lima, OH
Peoria, IL
Lawrence-Haverhill, MA-NH
Evansville, IN-KY

Indianapolis, IN
Springfield, IL
Wichita, KS
Decatur, IL
Terre Haute, IN
Anderson, IN

(continued on following page)

*This table is to be used in conjunction with Table 4-5 as discussed below pertaining to Group 4.
**Small SMSA

Hartford, CT
Minneapolis-St. Paul, MN-WI
San Jose, CA
[illegible], MI
Washington, DC-MD-VA
Norwalk, CT
Rochester, NY

[illegible] 4

Bristol, CT
Meriden, CT
New London-Norwich, CT-RI
Fitchburg-Leominster, MA

Baltimore, MD
Los Angeles, CA
Melbourne-Titusville-Cocoa Beach, FL
Daytona Beach, FL
Newark, NJ
Philadelphia, PA
Buffalo, NY

Appleton-Oshkosh, WI
Syracuse, NY
Racine, WI
Scranton, PA
Lorain-Elyria, OH
Worcester, MA
South Bend, IN
Binghamton, NY-PA

New Brunswick-Perth Amboy-Sayreville, NJ
Wilmington, DE-NJ-MD
Cleveland, OH
Detroit, MI
Trenton, NJ
Dayton, OH
Cincinnati, OH

Brockton, MA
Pittsfield, MA
Lowell, MA-NH
Kenosha, WI

Anaheim-Santa Ana-Garden Grove, CA
New Haven-West Haven, CT
Chicago, IL
Jersey City, NJ
Seattle-Everett, WA
San Francisco-Oakland, CA
Portland, OR-WA
New Britain, CT
Oxnard-Simi Valley-Ventura, CA
Nashville, TN
[illegible], CT

Green Bay, WI
La Crosse, WI
Dubuque, IA
Ogden, UT
Santa Cruz, CA
Hamilton-Middletown, OH
Albany-Schenectady-Troy, NY

[illegible]
Dallas, TX
Oklahoma City, OK
Raleigh, NC
Fort Worth, TX

Lexington, KY
Tallahassee, FL
Jacksonville, FL
Durham, NC
Memphis, TN-AR-MS

Des Moines, IA
Kansas City, MO-KS
Stamford, CT
Richmond, VA
Omaha, NE
Denver, CO
Columbus, OH
Houston, TX

[illegible] GROUPS

Fort Lauderdale-Hollywood, FL
Phoenix, AZ
Roanoke, VA
Sarasota, FL
Fall River, MA-RI
Albuquerque, NM

Eugene-Springfield, OR

Fargo-Moorhead, ND-MN
Lincoln, NE
Lafayette-West Lafayette, IN
Rochester, MN
Topeka, KS
Bloomington-Normal, IL

Canton, OH
Youngstown-Warren, OH
Pittsburgh, PA
Paterson-Clifton-Passaic, NJ
Utica-Rome, NY
Steubenville-Weirton, OH-WV
Johnstown, PA

Long Branch-Asbury Park, NJ
Atlantic City, NJ

Vineland-Millville-Bridgeton, NJ

Fort Smith, AR-OK
Lynchburg, VA
Asheville, NC

Las Vegas, NV

New York, NY

Madison, WI

Bryan-College Station, TX
Gainesville, FL
Columbia, MO
Austin, TX

[illegible]

[illegible], TN

Sioux Falls, SD
Reno, NV

Atlanta, GA

Baltimore, MD

Jackson, MI
Muskegon-Muskegon Heights, MI

Gary-Hammond, IN
Saginaw, MI
Muncie, IN
Bay City, MI

Flint, MI

Lansing-East Lansing, MI
Kalamazoo-Portage, MI

Ann Arbor, MI

Fayetteville, NC
Lawton, OK

Newport News-Hampton, VA
Norfolk-Virginia Beach-Portsmouth, VA
San Diego, CA

Champaign-Urbana-Rantoul, IL
Colorado Springs, CO

Tacoma, WA
Salinas-Seaside-Monterey, CA
Vallejo-Fairfield-Napa, CA
Great Falls, MT
Biloxi-Gulfport, MS

Miami, FL

Providence-Warwick-Pawtucket, RI-MA
Waterbury, CT

Springfield-Chicopee-Holyoke, MA-CT

Boston, MA

Bridgeport, CT
Macon, GA

Laredo, TX

McAllen-Pharr-Edinburg, TX

Brownsville-Harlingen-San Benito, TX

Table 6-2

Comparisons Between Selected Groups and Subgroups

	Stage II Factor			
	1	2	3	4
Group 1				
Subgroup 5				
New Orleans, LA	+	0	+	0
Portland, ME	+	0	0	0
Waco, TX	+	0	-	+
Little Rock-N. Little Rock, AR	++	+	-	+
Odessa, TX	++	+	-	0
Tulsa, OK	+	+	-	+
Baton Rouge, LA	++	+	-	+
Subgroup 6				
St. Joseph, MO	+	0	---	+
Sioux City, IA-NE	+	+	---	++
Billings, MT	++	++	--	0
Boise City, ID	++	+	--	+
Huntsville, AL	++	+	---	+
Springfield, MO	+	++	---	++
Amarillo, TX	++	+++	--	++
Group 4				
Subgroup 5				
Brockton, MA	--	0	+	--
Pittsfield, MA	--	0	+	--
Lowell, MA-NH	-	0	+	--
Kenosha, WI	--	0	0	---

Factor loadings are indicates as follows: .25 to .5 (+); .51 to 1.00 (++); 1.01+ (+++); -.25 to +.25 (0); -.25 to -.5 (-); -.51 to -1.00 (--); -1.01 or less (---).

Table 6-3

Characterization of SMSA Groups Through Stage II Factors

Stage II Factors	SMSA Groups 1	2	3	4	5
I. Low income, low level of manufacturing activity, low expenditures for sewerage	+++	+	-	-	0
II. Large SMSA with nonindustrial economic base experiencing rapid growth	+	-	--	-	++
III. Activity in minor industries, compact core, high expenditures for sanitation	-	-	--	+	+
IV. High employment, activity in services and trade	+	--	+	0	+++

Factor loadings are indicates as follows: .25 to .5 (+); .51 to 1.00 (++); 1.01+ (+++); -.25 to +.25 (0); -.25 to -.5 (-); -.51 to -1.00 (--); -1.01 or less (---).

Table 5-4

Characterization of SMSA Groups by Stage II Factors

	1	2	3	4
GROUP I				
Centroid	+++	+	-	+
Modal City: Little Rock, AR	++	+	-	+
GROUP II				
Centroid	+	-	-	--
Modal City: Lake Charles, LA	++	-	0	--
GROUP III				
Centroid	-	--	--	+
Modal City: Williamsport, PA	-	--	---	0
GROUP IV				
Centroid	-	-	+	0
Modal City: Albany-Schenectady-Troy, NY	--	0	0	0
GROUP V				
Centroid	0	++	+	+++
Modal City: Dallas, TX	+	++	+	+++
ALL CITIES				
Centroid	0	0	0	0
Modal City: Louisville, KY	0	-	0	0

Factor loadings are indicates as follows: .25 to .5 (+); .51 to 1.00 (++); 1.01+ (+++); -.25 to +.25 (0); -.25 to -.5 (-); -.51 to -1.00 (--); -1.01 or less (---).

rate of employment: unemployment in 1970 was at 3.3 percent (average for all SMSAs was 4.3 percent), and 34.7 percent of all housing units were built since 1960 (average for all SMSAs was 25.5 percent). Employment in manufacturing was below the national level: 20.1 percent as compared with 25.8 percent for all SMSAs in 1970. Other cities found near centroid of this group include Baton Rouge, LA; Corpus Christi, TX; and Montgomery, AL.

Group II includes cities which are closer to the centroid of all cities than Group I. Factor scores indicate these cities have high unemployment and are not active in services and trade. In addition, they generally have slightly lower than average income levels and economic activity and they may have experience less than average growth. Lake Charles, LA, the modal city for this group, is different from a hypothetical city at the centroid of the group, loading more heavily on Factor 1 and not at all on Factor 3 (see Table 6-4). A high factor 1 score is the result of a lower than average standard of living. Of all families in Lake Charles, 16.6 percent have incomes below the low income level and 4.7 percent of all housing units lack some or all plumbing facilities (national averages for SMSAs are 8.5 and 2.9 respectively). And, a large negative score on Factor 4 reflects the combined effect of slightly more than average activity in services and trade accompanied by very high unemployment (5.7 percent as opposed to the average 4.3 percent). Other cities near the centroid of the group are Spartanburg, SC, and Parkersburg, WV.

Group III has high negative scores on Factors 2 and 3. Cities in this group, then are expected to be small SMSAs (in area) with an industrial base and little recent growth. Williamsport, PA, the modal city for this group, includes only one county of moderate size, 42.6 percent of its employment is in manufacturing and population growth in the decade ending 1970 was only 3.6 percent, as compared with the national average of 16.6 percent for SMSAs. Other cities near the centroid of this group include Davenport-Rock Island-Moline, IA-IL; Evansville, IN-KY; and Lawrence-Haverhill, MA-NH.

Group IV is even closer to the overall centroid than Group II. Scores are moderately negative for Factors 1 and 2, moderately positive for Factor 3 and almost zero for factor 4 (see Table 6-3). Thus, these cities are expected to rank about average on the dimensions defined by the factor analysis. Table 6-4 shows, however, that the modal city, Albany-Schenectady-Troy, NY, has a more negative score than the centroid on Factor 1, the result of higher incomes and manufacturing activity, perhaps. Cities very similar to the modal city include Appleton-Oshkosh, WI, and New Britain, CT.

Group V has high scores on Factors 2 and 4, describing large SMSAs which are prosperous, as indicated by growth and high employment, and which are active in services and trade rather than manufacturing. Dallas, TX, has been designated as the modal city for this group and appears to fit the factor description. A great deal of caution should be exercised in dealing with the modal city and groups of cities, however, since the group includes a large number of cities for which several of the values were estimated. Most of these estimated values are for descriptors of the industry mix, which is important in the grouping of these cities. To avoid any incorrect conclusions from these groups or misuse of this classification, Table 6-5 should be used to identify biases created by the insertion of mean values as estimates.

Table [illegible]

Treatment of Missing Data

SMSA	Missing Values, Replaced by Mean Values: Number	Missing Values, Replaced by Mean Values: Codes
[illegible]		
Abilene, Texas		
Lafayette, Louisiana		
San Angelo, Texas	1	[illegible]
Midland, Texas		
Lubbock, Texas		
Tucson, Arizona		
Albany, Georgia		
Knoxville, Tennessee		
West Palm Beach-Boca Raton, Florida		
Tampa, Florida		
Monroe, Louisiana		
Orlando, Florida		
Sherman-Denison, Texas		
Corpus Christi, Texas		
Shreveport, Louisiana		
Macon, Georgia		
Texarkana, Texas-Arkansas	1	[illegible]
Wilmington, North Carolina		
Savannah, Georgia		
Tyler, Texas		
Montgomery, Alabama		
Jackson, Mississippi		
New Orleans, Louisiana		
Portland, Maine		
Waco, Texas	1	[illegible]
Little Rock-N. Little Rock, Arkansas		
Odessa, Texas		
Tulsa, Oklahoma		
Baton Rouge, Louisiana		
St. Joseph, Missouri	1	[illegible]
Sioux City, Iowa-Nebraska		
Billings, Montana		
Boise City, Idaho		
Huntsville, Alabama		
Springfield, Missouri		
Amarillo, Texas		

SMSA	Missing Values, Replaced by Mean Values: Number	Missing Values, Replaced by Mean Values: Code
[illegible]		
Huntington-Ashland, West Virginia-Kentucky-Ohio	1	[illegible]
Stockton, California		
Modesto, California		
Augusta, Georgia-South Carolina		
Pueblo, Colorado	1	[illegible]
Fresno, California		
Killeen-Temple, Texas	4	[illegible]
Lewiston-Auburn, Maine		
Pine Bluff, Arkansas	1	[illegible]
Spokane, Washington		
Owensboro, Kentucky		
Fort Myers, Florida	3	[illegible]
Yakima, Washington	3	[illegible]
Florence, Alabama	3	[illegible]
Santa Rosa, California		
Pensacola, Florida		
Riverside-San Bernardino-Ontario, California		
Provo-Orem, Utah	1	[illegible]
Charleston, South Carolina		
Salem, Oregon		
Duluth-Superior, Wisconsin-Minnesota		
Altoona, Pennsylvania		
Gadsden, Alabama		
Lake Charles, Louisiana		
Mobile, Alabama		
Bakersfield, California	2	[illegible]
Chattanooga, Tennessee-Georgia		
Lakeland-Winter Haven, Florida	5	[illegible]
Birmingham, Alabama		
Erie, Pennsylvania		
Wheeling, West Virginia-Ohio	1	[illegible]
Poughkeepsie, New York	3	[illegible]
Louisville, Kentucky-Indiana		
Richland-Kennewick, Washington	6	[illegible]
Springfield, Ohio		

Continued on following page

Table 6-5: continued

SMSA	Missing Values, Replaced by Mean Values Number	 Code
Parkersburg-Marietta, West Virginia-Ohio	12	[illegible], 9 manufacturing
Sacramento, California		
Gastonia, North Carolina		
Salt Lake City, Utah		
Petersburg-Colonial Heights-Hopewell, Virginia		
Charleston, West Virginia		
Wilkes Barre-Hazleton, Pennsylvania		
Beaumont-Port Arthur-Orange, Texas	2	[illegible]
Spartanburg, South Carolina		
New Bedford, Massachusetts		
Alexandria, Louisiana	13	[illegible], 9 manufacturing
El Paso, Texas		
Tuscaloosa, Alabama		
San Antonio, Texas		
Wichita Falls, Texas	1	[illegible]
Columbus, Georgia-Alabama		
Galveston-Texas City, Texas		
Manchester, New Hampshire	1	[illegible]
Santa Barbara-Santa Maria-Lompoc, California	1	[illegible]
Columbia, South Carolina		
GROUP 3		
Allentown-Bethlehem-Easton, Pennsylvania-New Jersey		
Harrisburg, Pennsylvania		
St. Louis, Missouri-Illinois		
Greenville, South Carolina	13	13 manufacturing
Reading, Pennsylvania		
York, Pennsylvania		
Lancaster, Pennsylvania		
Cedar Rapids, Iowa		
Waterloo-Cedar Falls, Iowa		
Fort Wayne, Indiana		
Toledo, Ohio-Michigan		
Rockford, Illinois		
Grand Rapids, Michigan		
Davenport-Rock Island, Moline, Iowa-Illinois		
Williamsport, Pennsylvania	3	[illegible]
Elmira, New York	3	[illegible]
Mansfield, Ohio		
Lima, Ohio		
Peoria, Illinois		
Lawrence-Haverhill, Massachusetts-New Hampshire		
Evansville, Indiana-Kentucky		
Indianapolis, Indiana		
Springfield, Illinois		
Wichita, Kansas	1	[illegible]
Decatur, Illinois		
Terre Haute, Indiana		
Anderson, Indiana		
Hartford, Connecticut		
Minneapolis-St. Paul, Minnesota-Wisconsin		
San Jose, California		
Milwaukee, Wisconsin		
Washington, D.C.-Maryland-Virginia		
Norwalk, Connecticut		
Rochester, New York		
GROUP 4		
Bristol, Connecticut	3	[illegible]
Meriden, Connecticut	4	[illegible]
New London-Norwich, Connecticut-Rhode Island		
Fitchburg-Leominster, Massachusetts		
Baltimore, Maryland	2	[illegible]
Los Angeles-Long Beach, California		
Melbourne-Titusville-Cocoa, Florida	4	[illegible]
Daytona Beach, Florida	3	[illegible]
Newark, New Jersey		
Philadelphia, Pennsylvania-New Jersey		
Buffalo, New York		
Appleton-Oshkosh, Wisconsin	1	[illegible]
Syracuse, New York		
Racine, Wisconsin		

continued on following page

Table [illegible]—continued

SMSA	Missing Values, Replaced by Mean Values	
	Number	Code
Scranton, Pennsylvania	13	13 manufacturing
Lorain-Elyria, Ohio		
Worcester, Massachusetts		
South Bend, Indiana		
Binghamton, New York-Pennsylvania		
New Brunswick-Perth Amboy-Sayreville, New Jersey	4	[illegible]
Wilmington, Delaware-New Jersey-Maryland		
Cleveland, Ohio		
Detroit, Michigan		
Trenton, New Jersey		
Dayton, Ohio		
Cincinnati, Ohio-Kentucky-Indiana		
Brockton, Massachusetts		
Pittsfield, Massachusetts	1	[illegible]
Lowell, Massachusetts-New Hampshire		
Kenosha, Wisconsin		
Anaheim-Santa Ana-Garden Grove, California		
New Haven-West Haven, Connecticut		
Chicago, Illinois	2	[illegible]
Jersey City, New Jersey		
Seattle-Everett, Washington		
San Francisco-Oakland, California		
Portland, Oregon-Washington	1	[illegible]
New Britain, Connecticut	2	[illegible]
Oxnard-Simi Valley-Ventura, California	3	[illegible]
Nashua, New Hampshire	6	[illegible]
Danbury, Connecticut	6	[illegible]
Green Bay, Wisconsin		
La Crosse, Wisconsin	2	[illegible]
Dubuque, Iowa		
Ogden, Utah	13	13 manufacturing
Santa Cruz, California	3	[illegible]
Hamilton-Middletown, Ohio	1	[illegible]
Albany-Schenectady-Troy, New York	1	[illegible]

SMSA	Missing Values, Replaced by Mean Values	
	Number	Code
GROUP [illegible]		
Charlotte, North Carolina	11	[illegible], 9 manufacturing
Dallas, Texas	13	[illegible], 9 manufacturing
Oklahoma City, Oklahoma		
Raleigh, North Carolina	13	13 manufacturing
Fort Worth, Texas	13	13 manufacturing
Lexington, Kentucky		
Tallahassee, Florida	3	[illegible]
Jacksonville, Florida		
Durham, North Carolina	13	13 manufacturing
Memphis, Tennessee-Arkansas-Mississippi		
Des Moines, Iowa		
Kansas City, Kansas-Missouri		
Stamford, Connecticut	1	[illegible]
Richmond, Virginia		
Omaha, Nebraska-Iowa		
Denver-Boulder, Colorado		
Columbus, Ohio		
Houston, Texas		
Fort Lauderdale-Hollywood, Florida		
Phoenix, Arizona		
Roanoke, Virginia		
Sarasota, Florida	3	[illegible]
Fall River, Massachusetts-Rhode Island		
Albuquerque, New Mexico		
Eugene-Springfield, Oregon		
Fargo-Moorhead, North Dakota-Minnesota		
Lincoln, Nebraska		
Lafayette-West Lafayette, Indiana		
Rochester, Minnesota		
Topeka, Kansas	1	[illegible]
Bloomington-Normal, Illinois		
Canton, Ohio		
Youngstown-Warren, Ohio		
Pittsburgh, Pennsylvania		
Paterson-Clifton-Passaic, New Jersey		
Utica-Rome, New York		
Steubenville-Weirton, Ohio-West Virginia	1	[illegible]
Johnstown, Pennsylvania		

continued on following page

Table 4-5: continued

SMSA	Missing Values, Replaced by Mean Values Number	Code
Flint, Michigan	2	[illegible], [illegible]
Lansing-East Lansing, Michigan	1	[illegible]
Kalamazoo-Portage, Michigan	2	[illegible], [illegible]
Ann Arbor, Michigan		
Fayetteville, North Carolina		
Lawton, Oklahoma	1	[illegible]
Newport News-Hampton, Virginia	1	[illegible]
Norfolk-Virginia Beach-Portsmouth, Virginia-North Carolina		
San Diego, California		
Champaign-Urbana-Rantoul, Illinois	1	[illegible]
Colorado Springs, Colorado		
Tacoma, Washington		
Salinas-Seaside-Monterey, California		
Vallejo-Fairfield-Napa, California	1	[illegible]
Great Falls, Montana	1	[illegible]
Biloxi-Gulfport, Mississippi		
Miami, Florida		
Providence-Warwick-Pawtucket, Rhode Island-Massachusetts		
Waterbury, Connecticut		
Springfield-Chicopee-Holyoke, Massachusetts-Connecticut		
Boston, Massachusetts		
Bridgeport, Connecticut		
Akron, Ohio		
Long Branch-Asbury Park, New Jersey	[illegible]	[illegible], 9 manufacturing
Atlantic City, New Jersey	7	7 manufacturing
Vineland-Millville-Bridgeton, New Jersey	1	[illegible]
Fort Smith, Arkansas-Oklahoma		
Lynchburg, Virginia		
Asheville, North Carolina		
Las Vegas, Nevada		

SMSA	Missing Values, Replaced by Mean Values Number	Code
New York, New York-New Jersey		
Madison, Wisconsin		
Bryan-College Station, Texas		
Gainesville, Florida		
Columbia, Missouri		
Austin, Texas		
Greensboro-Winston-Salem-High Point, North Carolina		
Nashville-Davidson, Tennessee		
Sioux Falls, South Dakota		
Reno, Nevada		
Atlanta, Georgia		
Battle Creek, Michigan	5	[illegible]
Jackson, Michigan		
Muskegon-Muskegon Heights, Michigan		
Gary-Hammond-East Chicago, Indiana		
Saginaw, Michigan		
Muncie, Indiana		
Bay City, Michigan		
Laredo, Texas	1	[illegible]
McAllen-Pharr-Edinburg, Texas		
Brownsville-Harlingen-San Benito, Texas		

6.4 Stability of the SMSA Groups

The groups described above were identified through a complex analysis of a very large data base. It is useful at this point to review some issues which relate to the manner in which these cities group. One of these issues, discussed briefly above, is the use of synthesized data where missing values were encountered. Other factors which influence the classification are the use of only four of the six Stage II factors, the equal weighting of all factors input to the cluster, and the clustering of all cities together rather than within strata distinguishing city size.

6.4.1 Number of Factors Used in the Classification --

As noted in Section 6.2, input to the cluster analysis was limited to 4 of the 6 factors available from Stage II. Together the 4 factors utilized explained 45 percent of the variance in Stage I factors as compared with 63 percent explained by 6 factors; this procedure was followed in order to simplify the resulting classification. As a precaution, however, a comparable clustering on 6 factors was performed to determine the effect on the clusters of eliminating the least important factors.

The major SMSA groups identified from 4 factors and from 6 factors are characterized in Table 6-6 by the centroidal factor scores for each group. Note from scores on Factors 1 to 4, that the groups are extrumely stable: Groups 1 through 5 seem to have a one-to-one correspondence even though the 6-factor classification yields an additional major group. The balance between groups, however, is very different; Table 6-7 provides some comparisons between the groupings. Elimination of the 2 secondary factors has the effect of consolidating small outlier groups by ignoring the more minor differences captured in the extra factors. The 4-factor classification yielded approximately half as many single-city groups, half as many groups of 2 to 5 cities, and 20 percent fewer groups of 6 to 10 cities. The distribution of cities between the major groups is more even, as well.

A more useful classification of cities has benn developed, then, at a cost equal to 18 percent of the variance in Stage I factors.

6.4.2 Weighting Factors --

The classification described in Section 6.3 was developed with equal weights attached to factors in the clustering process. In theory, a more accurate cluster could have been developed by attaching weights to the factor scores proportional to the amount of variance explained by that factor. The added expense of the weighting procedure was avoided in repeated iterations, however, a test on the final set of factors was performed to determine the sensitivity of SMSA groups to weighting the factors according to their explanatory power as experienced by their eigenvalues. The weights were as follows:

Factor 1: 1.91
Factor 2: 1.61
Factor 3: 1.59
Factor 4: 1.52

The number of cities which shifted from one group to another was slight.

Table 6-4

Centroids of Major Groups from Four and Six Factors

Group	SMSA	Number of Cities	Stage II Factors 1	2	3	4	5	6
Cluster on Four Factors								
1	0320 Amarillo, TX	36	+++	+	-	+		
2	3400 Huntington-Ashland, WV-KY-OH	43	+	-	-	--		
3	0400 Anderson, IN	27	-	--	--	+		
4	1170 Bristol, CT	48	-	-	+	0		
5	3360 Houston, TX	18	0	++	+	+++		
Cluster on Six Factors								
1	0320 Amarillo, TX	15	++	++	--	++	++	-
2	1440 Charleston, SC	59	+	0	0	-	0	-
3	0800 Bay City, MI	30	--	--	--	0	+	-
4	6780 Riverside-San Bernadino-Ontario, CA	25	0	-	+	0	+	++
5	1840 Columbus, Ohio	11	0	++	+	+++	-	--
6	1120 Boston, MA	14	--	-	+++	++	-	++

Key to factor scores: 0 = .25; +, - = .25-.65; ++, -- = .65-1.; and +++, --- = greater than 1.

Table 6-7

	4-Factor Classification	6-Factor Classification
Number of Cities in Major Groups:		
Group I	36	15
II	43	59
III	27	30
IV	48	25
V	18	11
VI		14
Number of Single-City Groups	17	33
Number of Cities in Groups of 1-to 5	61	92
Number of Cities in Groups of 6 to 11	87	108

Weighted factors should allow observations (SMSAs) which are outliers on a particular axis to group with other outliers on the same axis where the factor is of major importance, and to group with observations which are not outliers on that axis where the factor is of minor importance.

Adding factor weights to the clustering had the following effects:

- 5 outliers found a group;
- 2 cities, previously members of groups, became outliers;
- 7 cities moved from group 1 to group 2;
- 2 cities shifted to group 1 (from groups 3 and 5);
- 1 city shifted from group 2 to group 4

Because the effect of applying weights to the factors was relatively minor and the additional effort was not, the decision was made to ignore differences in explanatory power between factors except where the ratio between these is greater than 2.

6.4.3 SMSA Size --

To assess the bearing of city size on the classification, the set of 262 cities was divided into small (less than 200,000 population), medium (between 200,000 and 500,000 population) and large (greater than 500,000 population) SMSAs. A separate clustering, similar to that described for the entire sample, was performed within each of these strata. The resulting classifications are presented here as Tables 6-8, 6-9, and 6-10.

These separate classifications may be very useful where city size is of great importance. However, the general classifications appear to provide similar results. Note from Table 6-1 that small, medium, and large SMSAs are distributed throughout the general classification. And Table 6-11 shows that the groups formed within size strata have some similarities as well. For instance, small city group 1 is very similar to medium city group 1, and medium city group 3 is very similar to large city group 4. In any case, the four classifications are available and may be used where most appropriate.

Table 6-8

Classification of Small SMSAs
(population less than 200,000)

GROUP 1
Abilene, Texas
Billings, Montana
Boise City, Idaho
Odessa, Texas
San Angelo, Texas
Lafayette, Louisiana
Tyler, Texas
Sherman-Denison, Texas
Lubbock, Texas
Midland, Texas
Albany, Georgia
Springfield, Missouri
Amarillo, Texas

GROUP 2
Cedar Rapids, Iowa
Waterloo-Cedar Falls, Iowa
St. Joseph, Missouri
Sioux Falls, South Dakota
Springfield, Illinois
Decatur, Illinois
Fargo-Moorhead, North Dakota-Minnesota
Lincoln, Nebraska
Lafayette-West Lafayette, Indiana
Rochester, Minnesota
Topeka, Kansas
Bloomington-Normal, Illinois

GROUP 3
Galveston, Texas
Manchester, New Hampshire
Brockton, Massachusetts
Pittsfield, Massachusetts
Kenosha, Wisconsin
New Britain, Connecticut
Dubuque, Iowa
Santa Cruz, California
Ogden, Utah
Green Bay, Wisconsin
LaCrosse, Wisconsin
Nashua, New Hampshire
Danbury, Connecticut
Racine, Wisconsin
Anderson, Indiana
Terre Haute, Indiana
Elmira, New York

(continued on following page)

Table 6-8 (continued)

GROUP 3 (continued)
Springfield, Ohio
Parkersburg-Marietta, West Virginia-Ohio
Yakima, Washington
Florence, Alabama
Lake Charles, Louisiana
Provo-Orem, Utah
Salem, Oregon
Pueblo, Colorado
Killeen-Temple, Texas
Lewiston-Auburn, Maine
Pine Bluff, Arkansas
Owensboro, Kentucky
Fort Myers, Florida
Savannah, Georgia
Wilmington, North Carolina
Texarkana, Arkansas-Texas
Portland, Maine
Waco, Texas
Gastonia, North Carolina
Petersburg-Colonial Heights-Hopewell, Virginia
Richland-Kennewick, Washington
Spartanburg, North Carolina
New Bedford, Massachusetts
Alexandria, Louisiana

====================================

Wichita Falls, Texas
Tucson, Arizona

====================================

Durham, North Carolina
Tallahassee, Florida
Lexington, Kentucky

====================================

Roanoke, Virginia
Norwalk, Connecticut

====================================

Monroe, Louisiana

====================================

Sarasota, Florida
Fall River, Massachusetts-Rhode Island
Daytona Beach, Florida

====================================

Atlantic City, New Jersey
Wheeling, West Virginia-Ohio
Steubenville-Weirton, Ohio-West Virginia
Gadsden, Alabama
Altoona, Pennsylvania

====================================

(continued on following page)

Table 6-8 (continued)

Vineland-Millville-Bridgeton, New Jersey

=====================================

Fort Smith, Arkansas
Modesto, California
Lynchburg, Virginia
Asheville, North Carolina

=====================================

Bryan-College Station, Texas
Gainesville, Florida
Columbia, Missouri

=====================================

Champaign-Urbana-Rantoul, Illinois

=====================================

Great Falls, Montana
Biloxi-Gulfport, Mississippi

=====================================

Fitchburg-Leominster, Massachusetts
Meriden, Connecticut
Bristol, Connecticut

=====================================

Bay City, Michigan
Muncie, Indiana
Mansfield, Ohio
Williamsport, Pennsylvania
Lima, Ohio

=====================================

Jackson, Michigan
Muskegon-Muskegon Heights, Michigan

=====================================

Battle Creek, Michigan

=====================================

Sioux Falls, South Dakota
Reno, Nevada

=====================================

Laredo, Texas

=====================================

McAllen-Pharr-Edinburg, Texas

=====================================

Brownsville-Harlingen-San Benito, Texas

=====================================

Lawton, Oklahoma

=====================================

Table 6-9

Classification of Medium SMSAs

(population between 200,000 and [illegible],000)

GROUP 1

Albuquerque, New Mexico
Tuscon, Arizona
Huntsville, Alabama
Knoxville, Tennessee
West Palm Beach-Boca Raton, Florida
Orlando, Florida
Corpus Christi, Texas
Shreveport, Louisiana
Macon, Georgia
Montgomery, Alabama
Jackson, Mississippi
Little Rock-North Little Rock, Arkansas
Tulsa, Oklahoma
Baton Rouge, Louisiana

GROUP 2

Columbia, South Carolina
Spokane, Washington
Santa Barbara-Santa Maria-Lompoc, California
New London-Norwich, Connecticut
Oxnard-Simi Valley-Ventura, California
Lowell, Massachusetts-New Hampshire
New Haven-West Haven, Connecticut
Melbourne-Titusville-Cocoa, Florida

GROUP 3

Augusta, Georgia-South Carolina
Fresno, California
Pensacola, Florida
Santa Rosa, California
Charleston, South Carolina
Duluth-Superior, Minnesota-Wisconsin
Utica-Rome, New York
Johnstown, Pennsylvania

GROUP 4

Rockford, Illinois
Saginaw, Michigan
Bakersfield, California
Mobile, Alabama
Lakeland-Winter Haven, Florida
Charleston, West Virginia
Wilkes Barre-Hazelton, Pennsylvania
Beaumont-Port Arthur-Orange, Texas
Chattanooga, Tennessee-Georgia
Stockton, California

(continued on following page)

Table 6-5 (continued)

GROUP 4 (continued)

Huntington-Ashland, West Virginia-Kentucky-Ohio
Greenville, South Carolina
Harrisburg, Pennsylvania
Reading, Pennsylvania
York, Pennsylvania
Lancaster, Pennsylvania
Hamilton-Middletown, Ohio
Wichita, Kansas
Evansville, Indiana-Kentucky
Lawrence-Haverhill, Massachusetts-New Hampshire
Peoria, Illinois
Davenport-Rock Island-Moline, Iowa-Illinois
Canton, Ohio
Binghampton, New York-Pennsylvania
South Bend, Indiana
Worcester, Massachusetts
Lorain-Elyria, Ohio
Poughkeepsie, New York
Erie, Pennsylvania
Trenton, New Jersey
Wilmington, Delaware-New Jersey-Maryland
Scranton, Pennsylvania
Appleton-Oshkosh, Wisconsin

=====================================

Eugene-Springfield, Oregon

=====================================

El Paso, Texas
Columbus, Georgia-Alabama

=====================================

Long Branch-Asbury Park, New Jersey

=====================================

Charleston, North Carolina
Raleigh, North Carolina
Stamford, Connecticut
Des Moines, Iowa

=====================================

Madison, Wisconsin

=====================================

Las Vegas, Nevada

=====================================

Austin, Texas

=====================================

Newport News-Hampton, Virginia

=====================================

(continued on following page)

Table 6-9: (continued)

```
Fayetteville, North Carolina
=====================================

Salinas-Seaside-Monterey, California
Vallejo-Fairfield-Napa, California
Tacoma, Washington
Colorado Springs, Colorado
=====================================

Fort Wayne, Indiana
=====================================

Flint, Michigan
Lansing-East Lansing, Michigan
Kalamazoo-Portage, Montana
=====================================

Ann Arbor, Michigan
=====================================

Waterbury, Connecticut
======================================

Bridgeport, Connecticut
=====================================
```

Table 6-10

Classification of Large SMSAs

(population more than 500,000)

Akron, Ohio

====================================

GROUP 1
Youngstown, Ohio
Pittsburgh, Pennsylvania
Paterson-Clifton-Passaic, New Jersey

====================================

Grand Rapids, Michigan

====================================

Gary-Hammond-East Chicago, Indiana

====================================

GROUP 2
Columbus, Ohio
Denver-Boulder, Colorado
Omaha, Nebraska-Iowa
Richmond, Virginia
Kansas City, Missouri-Kansas
Houston, Texas
Fort Worth, Texas
Oklahoma City, Oklahoma
Dallas, Texas
Memphis, Tennessee-Arkansas-Indiana
Jacksonville, Florida

====================================

GROUP 3
Boston, Massachusetts
Newark, New Jersey
Philadelphia, Pennsylvania-New Jersey
Buffalo, New York

====================================

GROUP 4
Louisville, Kentucky-Indiana
Sacramento, California
Salt Lake City, Utah
New Orleans, Louisiana
Riverside-San Bernadino-Ontario, California

====================================

GROUP 5
Indianapolis, Indiana
Toledo, Ohio-Michigan
Baltimore, Maryland
Los Angeles-Long Beach, California
Allentown-Bethlehem-Easton, Pennsylvania-New Jersey
St. Louis, Missouri-Illinois
Rochester, New York
Syracuse, New York
New Brunswick-Perth-Amboy-Sayreville, New Jersey

(continued on following page)

Table 6-10 (continued)

Cleveland, Ohio
Detroit, Michigan
Dayton, Ohio
Cincinnati, Ohio-Kentucky-Indiana
Hartford, Connecticut
Minneapolis-St. Paul, Minnesota-Wisconsin
San Jose, California
Milwaukee, Wisconsin
Washington, D.C.-Maryland-Virginia
Anaheim-Santa Ana-Garden Grove, California
Chicago, Illinois
Jersey City, New Jersey
Seattle-Everett, Washington
San Francisco-Oakland, California
Portland, Oregon-Washington
Albany-Schenectady-Troy, New York

====================================

San Antonio, Texas

====================================

Miami, Florida

====================================

Ft. Lauderdale-Hollywood, Florida
Phoenix, Arizona

====================================

Tampa-St. Petersburg, Florida
Birmingham, Alabama

====================================

New York, New York-New Jersey

====================================

Greensboro-Winston-Salem-Highpoint, North Carolina

====================================

Nashville-Davidson, Tennessee

====================================

Atlanta, Georgia

====================================

San Diego, California
Norfolk-Virginia Beach-Portsmouth, Virginia-North Carolina

====================================

Springfield-Chicopee-Holyoke, Massachusetts-Connecticut

====================================

Providence-Warwick-Pawtucket, Rhode Island-Massachusetts

====================================

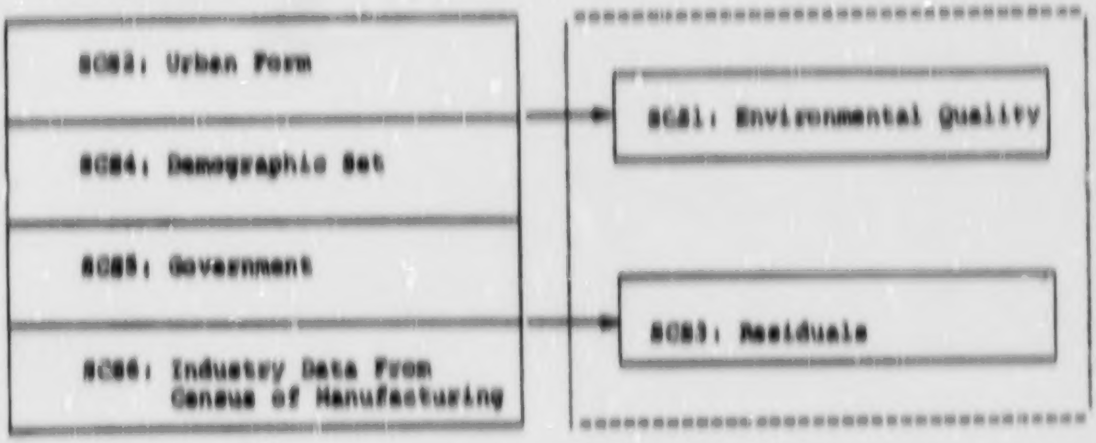
7.0 RELATIONSHIP OF THE BASIC SMSA GROUPS TO ENVIRONMENTAL QUALITY

The initial goal of this research was, as stated in Chapter 2, to group together metropolitan areas which are similar in terms of both environmental quality and factors which affect environmental quality. But, the basic classification described in Chapter 6 was developed without the benefit of information in SCS1 (environmental quality) or SCS3 (residual pollutants). The decision to exclude environmental measures from the initial classification is documented in section 5.2.1; the task at hand now is to integrate environmental data with the general city classification. (see Table 7-1).

The simplest method of utilizing information on the environment is to look at the variation in environmental measures between groups of SMSAs. Three approaches to such comparison have been executed: regression analysis, factor analysis, and t and F tests.

Table 7-1

Integration of Environmental Descriptors with SMSA Classification


SCS2: Urban Form
SCS4: Demographic Set
SCS5: Government
SCS6: Industry Data From Census of Manufacturing
SCS1: Environmental Quality
SCS3: Residuals

7.1 Regression Analysis

The regressions described here relate environmental quality variables to Stage I factors.* The hypothesis to be tested is that there is significant relationship between environmental measures and Stage I factors derived from, nonenvironmental data. The 8 environmental quality variables form the dependent variables, the 15 Stage I factors the independent variables.

---

*Environmental variables used are the 8 selected in Section 4.1.4; Stage I factors are the 15 which comprise the input to Stage II.

Preceding page blank

The statistics support the hypothesis. Table 7-2 summarizes the results of 8 separate regressions on the 15 Stage 1 factors. Note that, in general, the R^2 statistic is low; this seems reasonable in view of the fact that the factors are not expected to relate directly or linearly to the environmental variables (R^2 measures the proportion of variance in the environmental variable reflected by the factors). This statistic, however, is a better statistic for prediction than hypothesis testing since the proportion of variance will always increase with the number of independent variables.

The F-statistic, on the other hand, tests the hypothesis that there is a relationship between the dependent and independent measures by comparing the joint covariance to that resulting from a hypothetical set of random values. The F-statistics shown in Table 7-2 indicate at least a 99 percent probability that a significant relationship exists between each of the 8 environmental variables and the 15 factors.

Table 7-2

Regression Statistics

Dependent Variables	F*	R^2	Number of Factors with Significant t's**
DO	5.8	.28	6
BOD	3.2	.18	3
TURB	3.5	.20	3
HARD	2.6	.14	2
PHS	6.0	.30	7
NO3	2.8	.16	3
SUS	6.5	.29	3
PDI	3.1	.17	4

$^*F_{.99} = 2.1$

$^{**}t_{.95} = 1.96.$

Each of the 8 regressions is reviewed separately below and reported statistically in the series of Tables 7-3 through 7-10.

Dissolved Oxygen -- DO (Ambient)--The first point of interest is that DO rises with sewer expenditures (S5F2),* although it is negatively related to other sanitation (S5F4). Growing cities have, on the average, lower DO than older centralized cities (S2F1, S2F2, S4F1), and low income also is correlated with low stream DO. A portion of these effects may be related to the fact that northern cities naturally have higher DO because of lower temperatures. However, the general relationship is that sewers, higher incomes and slow growth all improve DO.

Biological Oxygen Demand -- BOD (Ambient)--BOD is most strongly related to S4F4, which shows cities with low auto use and few single family homes to have high stream BOD, possibly because of the increased population density and concentration. Both "heavy manufacturing" and "trade" increase BOD levels (S6F1, S6F2), partially because "trade" includes printing and publishing, along with services. Again, these relationships seem reasonable, and generally suggest that the higher density cities have higher BOD levels.

Turbidity -- TURB (Ambient)--Turbidity seems most closely related to industrial activity and the household sector; cities with high levels of turbidity show low levels of manufacturing, and high levels of petroleum, chemicals, and paper production. High turbidity is also related to sprawl and growth (probably because of erosion). Cities with young families and higher percentages of Blacks and crowded housing (i.e., the South, particularly Texas) show high turbidity. Thus, turbidity seems related to natural conditions, and to industrial and development practices.

Hardness -- HARD (Ambient)--Hardness seems to be correlated with sprawl, and the high percentage of single family homes and high auto use that go with it. Rubber processing and miscellaneous manufacturing also are found with increased water hardness. This is a difficult run to interpret since ambient water hardness is primarily determined by natural conditions; the relationships revealed are very weak.

Acidity -- PHS (Ambient)--Several variables seem related to stream pH:

More Alkaline:
Government Expenditures
Sewer Expenditures
Parks and Water Expenditures
Growth
Sprawl
Younger, more crowded populations
Central Employment
Heavy Manufacturing

---

*S5F2 indicates factor number 2 from SCS5.

Table 7-3

Regression Analysis: DO Against 15 Stage I Factors[1]

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	0.074	0.095	0.046	0.773
S5F2	0.214	0.103	0.130	2.077
S5F3	-0.020	0.095	-0.013	-0.214
S5F4	-0.220	0.104	-0.144	-2.119
S4F1	-0.384	0.117	-0.231	-3.285
S4F2	-0.033	0.109	-0.014	-0.310
S4F3	-0.670	0.128	-0.307	-5.214
S4F4	0.075	0.108	0.047	0.720
S2F1A	0.256	0.098	0.138	2.607
S2F2A	-0.375	0.109	-0.245	-3.441
S2F4A	0.041	0.093	0.026	0.437
S6F1	0.040	0.129	-0.024	-0.312
S6F2	0.197	0.115	0.120	1.715
S6F3	0.116	0.107	0.069	1.092
S6F4	0.054	0.100	0.033	0.542

Multiple R-square: .28; F Ratio: 5.8 (constant 8.2135)

Table 7-4

Regression Analysis: BOD Against 15 Stage I Factors

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	-1.040	1.470	-0.046	-0.707
S5F2	-0.074	1.501	-0.004	-0.050
S5F3	0.291	1.461	0.014	0.199
S5F4	-0.535	1.512	-0.026	-0.354
S4F1	2.328	1.745	0.103	1.334
S4F2	-0.113	1.654	-0.005	-0.068
S4F3	0.403	1.915	0.017	0.211
S4F4	-8.467	1.587	-0.383	-5.333
S2F1A	0.411	1.428	0.019	0.288
S2F2A	-0.115	1.595	-0.006	-0.072
S2F4A	-1.873	1.391	-0.090	-1.346
S6F1	4.350	1.990	0.187	2.186
S6F2	3.864	1.654	0.176	2.337
S6F3	0.536	1.529	0.024	0.351
S6F4	-1.531	1.440	-0.071	-1.063

Multiple R-square: .43; *F* Ratio: 3.2. (constant 14.4266)

Table 7-5

Regression Analysis: TURB Against 15 Stage I Factors

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	-0.411	0.444	-0.061	-0.926
S5F2	-0.546	0.478	-0.083	-1.141
S5F3	-0.715	0.478	-0.101	-1.496
S5F4	0.289	0.483	0.045	0.598
S4F1	0.657	0.561	0.090	1.171
S4F2	0.744	0.524	0.103	1.418
S4F3	0.963	0.605	0.131	1.593
S4F4	0.807	0.502	0.113	1.608
S2F1A	0.840	0.460	0.121	1.826
S2F2A	1.028	0.503	0.160	2.043
S2F4A	-0.064	0.435	-0.010	-0.147
S6F1	-1.204	0.622	-0.166	-1.937
S6F2	-0.672	0.531	-0.098	-1.265
S6F3	-0.667	0.481	-0.095	-1.386
S6F6	-1.427	0.454	-0.211	-3.142

Multiple R-square: 0.20; F Ratio: 3.5 (constant) 12.0588)

Table 7-6

Regression Analysis: HARD Against 15 Stage I Factors

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	14.729	18.203	0.052	0.809
S5F2	4.788	19.692	0.017	0.243
S5F3	17.441	18.201	0.064	0.958
S5F4	2.800	20.043	0.010	0.140
S4F1	21.083	22.373	0.073	0.942
S4F2	-18.191	20.875	-0.062	-0.871
S4F3	12.836	23.487	0.044	0.547
S4F4	68.557	21.903	0.219	3.130
S2F1A	35.771	18.581	0.127	1.925
S2F2A	20.761	21.041	0.076	0.988
S2F4A	30.414	18.081	0.109	1.682
S6F1	-0.793	25.251	-0.003	-0.031
S6F2	14.487	21.981	0.050	0.659
S6F3	-28.803	19.986	-0.096	-1.441
S6F4	48.647	19.234	0.167	2.529

Multiple R-square: 0.14; F Ratio: 2.6. (constant 239.8012)

Table 7-7

Regression Analysis: PHS Against 15 Stage I Factors

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	0.115	0.030	0.218	3.772
S5F2	0.059	0.033	0.115	1.802
S5F3	0.071	0.030	0.138	2.327
S5F4	-0.025	0.033	-0.048	-0.734
S4F1	0.077	0.037	0.143	2.084
S4F2	0.034	0.035	0.063	0.979
S4F3	-0.037	0.039	-0.067	-0.936
S4F4	0.083	0.035	0.150	2.397
S2F1A	0.124	0.031	0.238	4.049
S2F2A	-0.020	0.035	-0.039	-0.567
S2F4A	-0.010	0.029	-0.020	-0.337
S6F1	0.094	0.042	0.173	2.260
S6F2	0.009	0.037	0.017	0.260
S6F3	-0.070	0.033	-0.128	-2.122
S6F4	-0.052	0.032	-0.096	-1.631

Multiple R-square: 0.30; F Ratio: 6.8. (constant 7.4619)

Table 7-8

Regression Analysis: SO2 Against 15 Stage I Factors

Variable	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	-3.121	4.399	-0.047	-0.709
S5F2	2.559	4.800	0.040	0.533
S5F3	-5.230	4.545	-0.080	-1.151
S5F4	5.470	5.098	0.082	1.073
S4F1	-4.530	5.733	-0.061	-0.790
S4F2	3.737	5.347	0.052	0.699
S4F3	-2.698	6.640	-0.034	-0.406
S4F4	-1.544	4.858	-0.170	-2.376
S2F1A	1.797	4.364	0.028	0.412
S2F2A	3.659	5.322	0.055	0.687
S2F4A	0.485	4.517	0.007	0.107
S6F1	16.000	6.085	0.229	2.629
S6F2	-5.845	5.372	-0.084	-1.088
S6F3	-10.045	4.852	-0.147	-2.070
S6F4	4.276	4.598	0.063	0.930

Multiple R-square: 0.17; F Ratio: 2.8. (constant 76.6525)

Table 7-9

Regression Analysis: SUS Against 15 Stage I Factors

Variables	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	8.436	4.646	0.109	1.890
S5F2	-3.480	4.954	-0.049	-0.704
S5F3	14.689	4.617	0.190	3.181
S5F4	-4.184	5.131	-0.054	-0.815
S4F1	-3.702	5.552	-0.048	-0.667
S4F2	-7.878	5.216	-0.097	-1.510
S4F3	33.618	5.821	0.413	5.775
S4F4	6.802	5.055	0.083	1.345
S2F1A	2.760	4.500	0.036	0.613
S2F2A	-2.074	5.282	-0.027	-0.393
S2F4A	4.386	4.449	0.057	0.986
S6F1	3.838	6.323	0.046	0.607
S6F2	0.849	5.563	0.010	0.153
S6F3	-13.427	4.990	-0.163	-2.691
S6F4	-3.516	4.837	-0.043	00.727

Multiple R-square: 0.28; F Ratio 6.5. (constant 201.8078)

Table 7-10

Regression Analysis: PDI Against 15 Stage I Factors

Variables	Coefficient	Std.Error	Std. Reg. Coefficient	t
S5F1	0.158	0.334	0.031	0.474
S5F2	-0.228	0.371	-0.044	-0.613
S5F3	0.041	0.337	0.008	0.121
S5F4	0.766	0.374	0.150	2.046
S4F1	-0.779	0.404	-0.149	-1.927
S4F2	-0.403	0.376	-0.077	-1.070
S4F3	0.397	0.427	0.076	0.930
S4F4	-0.472	0.365	-0.090	-1.292
S2F1A	-1.056	0.353	-0.199	-2.992
S2F2A	1.105	0.397	0.213	2.784
S2F4A	0.161	0.343	0.030	0.468
S6F1	0.284	0.475	0.052	0.598
S6F2	0.276	0.404	0.051	0.682
S6F3	-0.994	0.365	-0.186	-2.723
S6F4	-0.069	0.355	-0.013	-0.196

Multiple R-square: 0.17; F Ratio: 3.1. (constant 3.2024)

More Acid: Larger Black Population
Mills
Miscellaneous Manufacturing
Petroleum and Chemicals

These effects can be related to land runoff, process effluents, and the natural conditions at the sites being analysed.

Sulfur Dioxide -- SO2--SO_2 is highest in cities with large amounts of heavy manufacturing, and is low in mill cities. It is also generally lower in sprawling cities than in centralized ones like New York, Boston, and Chicago. All these effects are as expected.

Particulates -- SUS--Particulates are especially high in low income cities with large amounts of crowded housing. The level also seems positively related to high government expenditures (especially on parks and water supply) and negatively related to mills. This is one of the least intuitive regressions of the set, and, therefore, one of the least conclusive.

Prevalence, Duration, Intensity Index -- PDI--PDI is generally higher for the larger, established cities, possibly because these cities have grown during a period when there was little concern over exceeding the assimilative capacity of our environment, therefore, the same types of activities might generate more pollution in these areas.

In summary, the DO, BOD, TURB, and SO2 regressions make good sense. The HARD and PHS regressions seem partially dominated by the natural environment masking human impact, and the PDI and SUS regressions revealed very little.

An additional set of regression was performed using Stage II factors rather than Stage I factors to ascertain whether a significant amount of explanatory power had been lost between the two stages. A comparison of F-statistics confirmed that the loss was significant, which is not surprising considering the reduction process involved. Despite this loss, however, the Stage II factors fill a useful function, as is confirmed by the discriminating power of the SMSA groups (see section 7.3 below).

7.2 Expanded Factor Analysis

The second approach to testing the relationship between environmental characteristics and more basic urban attributes was to include Stage I factors and variables from SCS1 and SCS3 in the Stage II factor analysis. First, the second stage analysis was repeated with each of the 8 SCS1 indicator variables, then with SCS3 factors. A close relationship between environmental and general attributes would be expected to cause the added measures to join Stage I factors from the other SCSs to form factors similar to those which resulted from the basic 15 factor set. If environmental attributes not closely aligned with other attributes, the added measures would cause the factors to be restructured to some extent--perhaps forming an entirely new Stage II factor.

When SCS1 indicator variables were added, the result was as expected--the indicator variables appended themselves directly to factors derived from the basic 15 factors. In no case was an additional factor developed. The variables tended to fall in Factors 1 and 5: 4 variables were most heavily loaded on Factor 1, 3 variables were most heavily loaded on Factor 5, and 1 aligned itself with Factor 2. Variables from the data base on environmental quality, then, correlate significantly ard meaningfully with the basic urban attributes described in Stage II analysis.

When SCS3 factors were included with the 15 factor set in factor analysis, the result was very different. SCS3 factors correlated poorly with factors from other SCSs. There was little basis for including these added measures in the factor analysis. Given the constraints in developing this data set (see Volume III), the lack of significant correlations between SCS3 and other Stage II factors is taken to mean the data utilized for SCS3 is inadequate for this type of analysis. This may also indicate that residuals data should not be used for comparing ambient environmental quality between regions.

7.3 Significance Tests between Groups

One of the tests planned for the analysis was to see whether the groups defined from the urban form, land use, government, and industry analyses also classified cities by environmental quality. In other words, were the groupings significantly different in the environmental quality variables. This is a standard hypothesis testing problem, to see whether the two sets of cities are drawn from the same population for the environmental variables.

In the simplest case--testing whether Group A and Group B have significantly different values for a single variable--a t-test is used with the null hypothesis indicating equal means for the two groups. For the combined sample sizes here, t-statistics of 1.96 to 2.05 are critical at the 95 percent level (one-tail), and the t-statistics are computed allowing unequal variance to avoid major errors. In other words, a t-statistic of approximately 2 indicates with 95 percent probability that the groups are different with respect to the variables on which the test was performed.

The first set of figures show the results for each variable. The groups tested were those described in Chapter 6 and we have characterized as

Group 1 (36 cities) -- low sanitation expenditures
low income
little manufacturing

Group 2 (46 cities) -- low employment
little trade and services

Group 3 (27 cities) -- industrial base
little recent growth
even population density
little miscellaneous manufacturing

Group 4 (48 cities) -- central; average cities

Group 5 (18 cities) -- large SMSA (in area)
prosperous
active in services and trade

Group 6 (87 cities) -- all others, some of which forms
smaller groups

Thus, *a priori* we would expect Groups 4 and 6 to be most typical and the other groups to show varied characteristics.

Table 7-11 shows the means of the 8 environmental indicators by group which vary considerably. The task is to determine which of these differences are significant, relative to the inherent variance of the measure. For each variable, then, each pair of groups was compared to generate t-statistics which indicate the magnitude of any difference in the means relative to the variance of the given variable. Tables 7-12 through 7-19 report these statistics.

Significant differences between groups were found for all but 1 variable--the PDI index--which has t's consistently less than 1.0. A good example of the other results is shown in Table 7-17, particulates. Groups 1, 2, and 5 are almost identical; so are Groups 3 and 4; but the 2 sets are significantly different. Group 6 is in between, almost significantly different from each of the other 5 groups (and, in fact, significantly different from the pooled 1-2-5 set and the pooled 3-4 sets).

Table 7-20 shows a summary table of the variables and city groups. As can be seen, each type of city is distinct, though some pairs are more similar than others. Groups 4 and 6 are the closest to typical, as expected.

Groups can also be compared in terms of general environmental quality by performing F-tests between the groups using all 8 indicator variables together. The null hypothesis is that no group has its own characteristics, and that a pair can be combined for their environmental quality effects.

We are interested in comparing the means of a set of variables between two samples, for example, Groups 1 and 2. The F-statistic is used, and is, really, just a generalization of the t-statistic.

For this analysis, PDI was dropped because we know that it cannot be distinguished, and the F-statistic analyses whether the other 7 environmental variables are, in general, significantly different among the pairs of groups. A further generalization, to see whether the distribution of all variables over all groups was significant, was not performed because it would have added little. This was clearly true from the pairwise results.

Table 7-11

Mean Values of Environmental Indicators by Group

	GROUP					
	1	2	3	4	5	6
DO	7.67	8.08	9.40	6.46	7.62	8.10
BOD	14.28	9.92	19.73	14.35	19.28	15.56
HARD	336.50	149.50	222.60	360.40	257.50	241.70
PHS	7.39	7.29	7.72	7.48	7.52	7.45
TURB	14.40	13.30	11.14	8.96	15.21	11.21
SUS	217.00	223.00	181.20	180.70	224.30	198.90
SO2	37.60	77.00	93.60	103.00	98.50	74.70
PDI	2.57	2.84	3.15	3.75	2.40	3.39

Table 7-12

Comparison of SMSA Groups Using DO as the Measure
(t-statistic)

Group	1	2	3	4	5	6
1	X					
2	1.07	X				
3	5.13	3.83	X			
4	2.35	1.11	3.21	X		
5	0.10	1.02	4.33	2.04	X	
6	1.38	0.07	4.91	1.36	1.22	X

Similar Groups	DO Level
5=1	Low
2=6	Medium
4	High
3	Very High

Table 7-13

Comparison of SMSA Groups Using BOD as the Measure
(t-statistic)

Group	1	2	3	4	5	6
1	X					
2	0.76	X				
3	0.85	2.28	X			
4	0.01	1.29	1.22	X		
5	0.79	2.21	0.09	1.13	X	
6	0.23	1.59	0.93	0.33	0.84	X

Similar Groups	BOD Level
2	Low
1=4=6	Medium
5=3	High

Table 7-14

Comparison of SMSA Groups Using HARD as the Measure
(t-statistics)

Group		1	2	3	4	5	6
Group	1	X					
	2	2.50	X				
	3	1.58	2.04	X			
	4	0.86	1.80	0.65	X		
	5	0.70	1.15	0.38	0.03	X	
	6	1.30	2.50	0.61	0.32	0.17	X

Similar Groups	HARD Level
2	Low
3-4-5-6	Medium
1	High

Table 7-15

Comparison of SMSA Groups Using PHS as the Measure
(t-statistics)

Group		1	2	3	4	5	6
Group	1	X					
	2	0.94	X				
	3	2.89	3.75	X			
	4	0.81	1.78	2.17	X		
	5	1.06	1.92	1.66	0.33	X	
	6	0.67	1.71	2.53	0.22	0.55	X

Similar Groups	PHS Level
1-2-4-5-6	Medium
3	High

Table 7-16

Comparison of SMSA Groups Using TURB as the Measure

(t-statistics)

Group	1	2	3	4	5	6
1	X					
2	0.63	X				
3	1.73	1.23	X			
4	3.57	3.20	1.42	X		
5	0.40	1.01	2.01	3.68	X	
6	2.10	1.55	0.05	2.15	2.36	X

Similar Groups	TURB Level
4	Low
3-6	Medium
2-1-5	High

Table 7-17

Comparison of SMSA Groups Using SUS as the Measure

(t-statistics)

Group	1	2	3	4	5	6
1	X					
2	0.31	X				
3	2.21	2.65	X			
4	2.38	2.87	0.03	X		
5	0.34	0.11	2.04	2.13	X	
6	1.19	1.60	1.76	1.33	1.24	X

Similar Groups	SUS Level
1-2-5	High
6	Medium
3-4	Low

Table 7-18

Comparison of SMSA Groups Using SO2 as the Measure
(t-statistics)

		Group 1	2	3	4	5	6
Group	1	X					
	2	3.35	X				
	3	5.59	1.28	X			
	4	5.45	1.78	0.71	X		
	5	2.24	0.76	0.18	0.15	X	
	6	4.03	0.18	1.76	2.24	0.87	X

Similar Groups	SO2 Level
3-4-5	High
2-6	Medium
1	Low

Table 7-19

Comparison of SMSA Groups Using PDI as the Measure
(t-statistics)

		Group 1	2	3	4	5	6
Group	1	X					
	2	0.38	X				
	3	0.56	0.30	X			
	4	1.16	0.89	0.48	X		
	5	0.24	0.62	0.72	1.32	X	
	6	0.93	0.63	0.21	0.31	1.17	X

Groups 1 through 6 do not show significant differences in the PDI rating.

Table 7-21 reports the values of the between group comparisons, and the relevant critical F-statistics for 7 degrees of freedom. Underlined numbers show pairs of groups where the probability of their being distinct is less than .95. If the probability was zero, or close to it, we could pool the two groups and conclude the cities in each show similar environmental quality. In every case, however, F-statistics indicated with at least 60 percent probability that the groups were significantly different given the 8 environmental indicators.

The groups of interest, then, all show significantly different levels of environmental quality. Group 6--the grabbag of "all other cities" is not relevant, of course, and it might be possible to pool the cities in groups 3 and 5. As expected, Group 6 appeared in the middle set in all cases, and had consistently higher unit variation. These results indicate that the groups do significantly better than a random set in classifying cities by environmental quality. They may not be the best possible groupings for that purpose (since those would be achieved by just clustering the environmental variables) but they are significant; which means that the factor and cluster procedures are effective.

Table 7-20

Levels of Environmental Quality by SMSA Group

Variable	Group 1	2	3	4	5	6
DO	L	M	VH	H	L	M
HARD	H	L	M	M	M	M
BOD	M	L	H	M	H	M
TURB	H	H	M	L	H	M
SO2	L	M	H	H	H	M
SUS	H	H	L	L	H	M
PHS	M	M	H	M	M	M
PDI	M	M	M	M	M	M

L--low; M--medium; H-high; VH--very high

Table 7-21

Comparison of City Groups Over all Environmental Variables

(F-statistics)

Group	1	2	3	4	5	6
1	X					
2	$2.77_{7,61}$	X				
3	$8.92_{7,46}$	5.13	X			
4	$5.36_{7,63}$	4.57	2.06	X		
5	$2.36_{7,39}$	1.31	4.72	3.26	X	
6	$3.96_{7,99}$	1.82	3.47	1.50	1.24	X

$F_{.95}$ (7, 40) = 2.25

$F_{.95}$ (7, 60) = 2.17

$F_{.95}$ (7,120) = 2.09

Groups 1, 3, and 4 are distinct over the seven variables. Groups 2 and 5 could probably be pooled. Group 6 is central but high variance, could be pooled but probably should not be.

8.0 CLASSIFICATION OF SMSAs FOR ALTERNATIVE PURPOSES

The data base and classification process developed for this research can be useful for a much broader range of research problems than the general question approached in the preceding chapters. The set of variables, the set of SMSAs, and the factor analysis technique (one or two stage) may be selected to suit alternative research hypotheses (see section 8.1). One example of other potential uses is a classification performed for a problem proposed by the Energy Research Development Agency (ERDA) with a view toward alternative transportation policies in urban areas, particularly electric car implementation. This is described in detail in Section 8.2.

8.1 Potential Applications

The data base and the methodology developed during this research project may be utilized for alternative research purposes in three main categories. These are: the direct use of the data base, case study/demonstration project site selection, and the extrapolation of case study results to group results.

The direct use of the data base does not merit extensive discussion. Clearly, researchers requiring the information available from this data base should not duplicate data collection efforts, particularly since some of the information has been obtained from unpublished sources. A comprehensive description of the data sources, strengths and weaknesses, as well as a data listing, may be found in Volume III.

Case study or demonstration project sites may be selected for a variety of research purposes using the data base and methodology developed in this project. The major constraint is that the research hypotheses to be tested must be capable of being framed in terms of the variables included in the data base. Indeed, a subset of the data base may be adequate for that purpose. For example, if a program analyst were interested in studying the effects of an antipoverty program, environmental quality and urban form descriptors would be of peripheral interest for his research purposes. Alternatively, if air quality maintenance programs represented the focal point of the inquiry, then water quality descriptors, and some of the income variables, may not be relevant. The first step in developing appropriate city groupings is the specification of the relevant data set to be used for factor analysis and clustering. Secondly, the universe of SMSAs may be limited to fit the requirements of the research project. The program being tested may apply to a limited geographical area such as the South, or may be relevant for cities in a certain size category only. It is possible to identify cities with certain attributes to be excluded/included from the analysis. Once the data set and the universe of SMSAs is delimited, factor analysis is performed. If the data set is composed of a limited number of variables, and if the research hypotheses are relatively simple and well defined, a simple one stage factor analysis may be adequate. A more complicated research design may call for a two stage factor analysis, as done for the general environmental analysis. Clustering is then performed on the basis of the first or second stage factors obtained, and representative cities are selected for each of the city groups. The case studies, or demonstration projects, should be sited at these representative cities to best ensure that their results can be generalized to the other SMSAs in the group.

In some cases, it may not be possible to perform a case study at an ideal location; in the appropriate representative city because of costs, data limitations, or simply due to lack of cooperation. Alternative sites may then be chosen. Further, the site selection process may be random, based on less rational criteria than the ones described here. It is appropriate to ask whether the results of such studies may be extrapolated to other sites. The data base and methodology developed here may be used for this prupose as well. The data base and the universe of SMSAs must be delimited first, and second, factor analysis and clustering must be performed as described above. The site of the case studies/demonstration projects must be identified relative to the city groups with the results being capable of generalization to the other cities in the groups. If the sites do not fall into any of the groups, then the study results may not easily be generalized to other cities. A large number of case studies have been performed in outlier cities such as New York, not because of a random site selection process but because the problem areas are frequently outliers. It is possible to analyze the data for such SMSAs, and to determine the axes or variables with extreme observations, which are in fact responsible for the outlying position of the SMSA. If these variables are not crucial to the analysis, the SMSA may be grouped with other cities in terms of the remaining variables. Study results may be generalized to this group--although at lower levels of confidence.

8.2 Transportation Analysis Application

In the context of this project, a specific city classification scheme was developed in response to a problem proposed by the Energy Research Development Agency. ERDA is concerned with the potential for energy conservation which may be achieved through alternative transportation policies, in urban areas, in particular, through the use of electric cars. The development of an appropriate urban classification scheme and the identification of SMSAs for performing case studies and/or siting demonstration projects the objective for this task.

Given the more limited scope of this classification, the factor analysis was performed in a single stage. A set of 13 variables was chosen from the assembled data base to reflect attributes of an urban area which are important in urban transportation analysis (see Table 8-1).

Table 8-1

Variables Selected for the Transportation Classification

Variable	Description
1. H1U	Percent of families in single-unit housing
2. TD	Percent of workers commuting as auto drivers
3. AUTO	Percent of households with one or more cars
4. CENE	Percent of SMSA employment in central city
5. PCH	Percent population change, 1960-1970
6. EFF	Percent of women 18 or older who are employed
7. YM	Median household income
8. PBL	Percent of population which is black
9. PPH	Persons/housing unit
10. SQMU	Square miles/person, urbanized area
11. RAD	Count of radial highways
12. CIR	Count of circumferential highways
13. VMTP	Vehicle miles travelled/capita-day

The transportation analysis is particularly interesting because it has been performed within three city size strata as well as for the full sample of 262 SMSAs. For reference purposes, the means of variables by city size are presented as Table 8-2.

In four separate factor analyses performed, a total of 12 distinct factors were generated. The sorted factor loadings are shown in Table 8-3 through Table 8-6, and a summary listing of the factors in Table 8-7. Each of these factors will be interpreted in turn.

As can be seen from Table 8-7, a few factors dominate all the runs, and the others very somewhat. This is not surprising--it simply means that certain attributes are more prominent among cities of a given class. For example, VMTP (daily vehicle-miles per capita) appears in three distinct factors, one suggesting that VMTP is higher in SMSAs which have low density and less central activity; another relating VMTP inversely with household size; and the third suggesting that blacks have lower levels of VMTP than the rest of the population. Race, in this case, is probably acting as a surrogate for city size, given the appearance of this factor in the analysis of small cities. These effects are interpretable and fit observed evidence from other studies, but each seems to be important within a particular class of cities.

Factor 1: This factor includes auto ownership, auto commuting behavior, and, to a lesser extent, single unit housing and population growth. All 4 runs show evidence of this effect, and in 3 it is the strongest factor.

Factor 2: The central relationship here is a positive loading on PBL (black population) and a negative one on YM (median income), showing the expected relationship between race and income.

Factor 3: If female employment is often an alternative to child-raising, smaller households would be expected in cities with high female employment. This tendency is shown to be particularly strong in large cities, but also appears in medium cities (factor 2) and small cities (factor 1) at lower loadings.

Factor 4: Since the radials and circumferentials were counts of the number of highways, they should be larger for bigger cities, and this is shown by their mean values for city groups. They should also be related in the analysis, since communities with one may be expected to have the other. The large and small city runs show this effect strongly.

Factor 5: The large city runs show a relationship between density and VMTP, with lower densities leading to higher VMTP; the run for all cities suggests that VMTP is negatively related to the concentration of activity in the central city, so that travel rises when employment moves to the suburbs. The small cities also show a relationship of this type.

Factor 6: As mentioned above, VMTP and persons/household are related for the medium SMSAs, with a secondary loading similar to Factor 5.

Table 8-2

Means for Variables by City Size Group

Variable	Units	Small*	Medium**	Large***	All Cities
Number of SMSAs		110	87	65	262
Families in single-unit housing	percent	75	75	68	73
Workers commuting as auto drivers	percent	71.3	71.6	68.8	70.6
Households with one or more cars	percent	86.0	86.7	83.8	85.7
SMSA employment in central city	percent	57.4	47.0	45.5	51.0
Population change, 1960-1970	percent	15.7	20.7	21.3	18.7
Women 18 or older who are employed	percent	28.0	27.5	29.0	28.1
Median household income	$000	8995	9697	10484	9597
Population which is black	percent	8.2	9.9	11.6	9.6
Number of persons per housing unit	pers./hsehld.	3.1	3.2	3.1	3.1
Square miles per person—urbanized area	sq.mi./person	46.0	39.9	36.6	42.6
Count of radial highways	number	6.2	7.7	12.4	8.3
Count of circumferential highways	number	.18	.28	.59	.31
Vehicle miles travelled/capita day	veh.mi/per./day	9.5	9.9	10.1	9.8

*Small—population less than 200,000 in SMSA
**Medium—population 200,000 to 500,000 in SMSA
***Large—population greater than 500,000 in SMSA

Table 8-3

Factor Analysis for Large Cities

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Workers commuting as auto drivers	++				
Households with 1 or more autos	++				
Families in single-unit housing	++		+		
Population change, 1960-70	++		-		
SMSA employment--central city		++			
Median household income		--		+	
Black population		+		+	
Persons/housing unit			++		
Women 18 or older who are employed			--	+	
Number of radial highways				++	
Number of circumferential highways				+	--
Square miles/person--urbanized area					+
Vehicle miles travelled/capita day					+

Table 8-4

Factor Analysis for Medium Cities

Variable	Factor 1	Factor 2	Factor 7	Factor 8	Factor 9
Households with one or more cars	++				
Population change, 1960-70	++		+		-
Workers commuting as auto drivers	++	+	-		
Families in single-unit housing		++			
Black population		+	+	+	
Median household income	+	-			
Square miles/person--urbanized area		+			+
Number of circumferential roadways			++		
Women 18 or older who are employed		-	++		
Persons/housing unit		+		++	
Vehicle miles travelled/capita day				--	
SMSA employment in central city				+	++
Number of radial highways					++

Table 8-5

Factor Analysis for Small Cities

Variable	Factor 9	Factor 1	Factor 4	Factor 11	Factor 12	Factor 10
Women 16 or older who are employed	++		+			
Median household income	++	+				-
Families in single-unit housing	-	+				
Households with one or more cars		++				
Workers commuting as auto drivers		++				
Number of radial highways			++			
Number of circumferential highways		+	+			-
Persons/housing unit	-		-	++		
Population change, 1960-70	+		-	-		
SMSA employment in central city	+		+	+		
Vehicle miles travelled/capita-day				-	++	
Black population				-	--	
Square miles/person--urbanized area						++

Table 8-6

Factor Analysis for All Cities

Variable	Factor 1	Factor 3	Factor 4	Factor 2	Factor 5
Households with one or more cars	++				
Workers commuting as auto drivers	++				
Population change, 1960-70	+				
Women 16 or older who are employed		--	+		+
Persons/housing unit		++			
Number of radial highways			++		
Number of circumferential highways			++		
Median household income		-	+	--	
Black population			+	++	
Square miles/person, urbanized area				+	
SMSA employment in central city					++
Vehicle miles travelled/capita-day		-			-
Families in single-unit housing	+	+		+	

Table F-7

Factors Relevant to Urban Transportation Analysis

Factor	City Size for Which Factor is Important Large	 Medium	 Small	 All
1. Auto use	X	X	X	X
2. Income/race	X	X		X
3. Female employment/ occupancy	X			X
4. Highways	X		X	X
5. VMTP/density	X			X
6. VMTP/occupancy		X		
7. Circumferentials/ employment		X		
8. Centrality/radials		X		
9. Employment/income			X	
10. Density			X	
11. Growth/centrality			X	
12. Race/VMTP			X	

Factor 7: This factor suggests a relation between higher levels of black population, female employment, and circumferentials only for SMSAs in the 200,000 to 500,000 range. This factor appears to be an indicator of city size more than anything else.

Factor 8: The final factor for medium cities relates the number of radials to the level of central city activity, an effect which could easily be masked in larger SMSAs because of multiple centers and in the smaller ones because of the ways SMSAs are defined. It does appear in a weaker form in the small city runs.

Factor 9: Factor 9, for small cities, shows a positive relationship between female employment and income, but a negative correlation to household size (see factor 3) and to single unit housing. The older, high-income northern cities (e.g., most of New England) are positive on the factor; the poor southern cities with large numbers of group quarters are negative (e.g., Laredo). Thus, the income effects predominate and the housing indicates family structure and life-styles.

Factor 10: This factor separates out the density effect, which is defined by the urbanized area of the SMSA. Since a number of the smaller SMSAs do not have urbanized areas defined, and a number of others are completely urbanized, the variable is a weak measure at best, and its separation is encouraging.

Factor 11: This factor shows the negative relationship between growth (1960-1970) and central activity (1970). Where growth occurred, it was primarily suburban and reduced central city activity. The older established manufacturing centers show less growth.

Factor 12: This factor shows a negative relationship between VMTP and the black population, not unexpected when income, auto ownership, and travel characteristics are considered.

The second step in the process was to choose factors to be used in classifying SMSAs, in each case. For the full sample and for large cities, the factors related to numbers 1, 2, 4 and 5 on the above list were chosen. For the medium cities, numbers 1, 3, 6, and 8 were included, and for the small cities, numbers 1, 4, 9 and 12. The number of factors was limited to 4 in each case to insure a reasonable amount of grouping.

We have interpreted the clusters only for the three runs of different city sizes, since there were some differences between the factors and since policy use of the results is expected to be contingent on city size.

For the large cities, there are two major groups (of 20 and 11 cities) and groups of 6, 4, 4 and 3 in the 64 cities. Table 8-8 describes these groups in the four-dimensional factor space.

Table 8-8

Classification of Large Cities

Group	Number of Cities in Group	Modal City	Centroid Location (by Factor Scores) Factor 1	Factor 2	Factor 4	Factor 5
1	11	Providence	-.49	-.81	-.22	.86
2	20	Louisville	.04	-.16	.17	-.48
3	6	Portland, OR	.44	-.13	-1.37	.40
4	3	Oklahoma City	.93	.69	-.22	1.07
5	4	Toledo	.25	.61	1.27	.90
6	4	Norfolk, VA	-.19	2.29	.10	.42
TOTAL	48	NOTE: There are 17 outliers in this classification.				

Modal cities have been identified using the measure of minimum distance from group centroid. About a quarter of the cities count as outliers on these 4 factors, which is reasonable given the wide divergence in character shown by the largest cities.

Group 1 includes cities which are low on Factors 1 and 2 and high on Factor 5. These cities have somewhat low auto ownership and use, show zero growth, relatively high income, few highways, but also low density. The dominant effects are low density, high income, and low central activity.

Group 2 is the central group. Its strongest coordinate is one-half standard deviation from the centroid of all cities, and its cities are the nearest to average of the whole population. Group 3 is strongly negative on Factor 4, showing cities which are low in number of highways.

Group 4 is high on auto ownership and use, as might be expected for Dallas, Fort Worth, and Oklahoma City. Group 5 is the inverse of Group 4, with many radial highways, and Group 6 is low income with strong central activity. The major transit cities (Boston, Chicago, New York) are distinct enough to be outliers.

The 84 medium cities also show 2 major groups, with 6 significant smaller groups and only 9 cities are outliers. Modal cities and centroids are shown in Table 8-9. As before, Group 2 is the largest and most central. The interpretations of the city types is a simple matter, given the centroids and the factor descriptions above. For example, Group 7 is very high on auto use, and has low central city activity and few radials. Group 1, the other major group, is above average in central city activity and radials, somewhat above average in VMTP, and below average in household size.

Table 8-9

Classification of Medium Cities

Group	Number of Cities in Group	Model City	Centroid Location (by Factor Scores) Factor 1	Factor 2	Factor 6	Factor 8
1	21	Little Rock	-.13	.34	-.66	.64
2	29	Tacoma	.41	-.50	.18	-.39
3	5	Harrisburg	-1.01	-.25	1.50	-1.06
4	3	West Palm Beach	.60	.40	-1.65	-.79
5	4	Worcester	-.85	-1.81	.12	1.22
6	7	Newport	-.79	.82	1.49	.63
7	3	Melbourne	2.60	-.24	.74	-1.23
8	3	Augusta	-1.22	1.14	1.73	-1.69
Total	75	NOTE: There are 9 outliers in this classification				

Within the 107 small cities, there are 5 major groups, 3 smaller groups, and about 10 percent unclassified cities. As before, the largest cluster includes the most typical cities--in this case, Group 5. Of the other major groups, Group 2 shows a large number of highways, some central city concentration, high VMTP and a low Black population, on the average. Group 3 includes the high income cities with high female employment and low percentage of single unit housing--in this case, primarily the older developed cities. Group 4 is largely a racial factor, showing cities with high black population, low incomes, and low VMTP. Finally, Group 6 is the cities which have low incomes and low auto ownership and use.

Table 8-10

Classification of Small Cities

Group	Number of Cities in Group	Model City	Centroid Location (by Factor Scores) Factor 9	Factor 1	Factor 4	Factor 12
1	4	Abilene	-.16	1.91	1.22	.18
2	11	St. Joseph	.07	.02	1.44	.79
3	13	Lincoln	1.60	-.14	.36	.11
4	8	Spartanburg	-.58	-.27	.77	-1.08
5	44	Sarasota	-.11	.38	-.41	-.04
6	9	Parkersburg	-1.24	-.06	-.86	.70
7	3	Lewiston, ME	1.40	-2.34	.42	-.02
8	3	Albany, GA	.46	-1.26	-.42	-2.20
TOTAL	95	NOTE: There are 12 outliers in this classification.				

The SMSA groups are tabulated in Tables 8-11 through 8-13.

Table 8-12

Classification of Large Cities

GROUP 1

Akron, Ohio
Youngstown-Warren, Ohio
Dayton, Ohio
Newark, New Jersey
Milwaukee, Wisconsin
Providence-Warwick-Pawtucket, Rhode Island-Massachusetts*
Cleveland, Ohio
Rochester, New York
Paterson-Clifton-Passaic, New Jersey
Springfield-Chicopee-Holyoke, Massachusetts-Connecticut
Pittsburgh, Pennsylvania

GROUP 2

Atlanta, Georgia
Columbus, Ohio
Denver, Colorado
St. Louis, Missouri-Illinois
Detroit, Michigan
Baltimore, Maryland
Gary-Hammond, Indiana
Kansas City, Missouri-Kansas
Louisville, Kentucky-Indiana
Omaha, Nebraska
Los Angeles-Long Beach, California
Philadelphia, Pennsylvania-New Jersey
San Francisco-Oakland, California
Cincinnati, Ohio-Kentucky-Indiana
Allentown-Bethlehem-Easton, Pennsylvania-New Jersey
Seattle-Everett, Washington
Grand Rapids, Michigan
Miami, Florida
Syracuse, New York
Albany-Schenectady-Troy, New York

GROUP 3

Tampa-St. Petersburg, Florida
Riverside-San Bernardino-Ontario, California
San Diego, California
Sacramento, California
Salt Lake City, Utah
Portland, Oregon-Washington*

GROUP 4

Dallas, Texas
Oklahoma City, Oklahoma*
Fort Worth, Texas

GROUP 5

Indianapolis, Indiana
Toledo, Ohio-Michigan*
Richmond, Virginia
Greensboro-Winston-Salem-High Point, North Carolina

GROUP 6

Jacksonville, Florida
Memphis, Tennessee-Arkansas-Mississippi
Norfolk-Virginia Beach-Portsmouth, Virginia-North Carolina*
Birmingham, Alabama

UNCLASSIFIED

Nashville-Davidson, Tennessee
New Brunswick-Perth Amboy-Sayreville, New Jersey
Minneapolis-St. Paul, Minnesota-Wisconsin
Washington, D.C.-Maryland-Virginia
Chicago, Illinois
Boston, Massachusetts
San Antonio, Texas
Phoenix, Arizona
Ft. Lauderdale-Hollywood, Florida
Buffalo, New York
San Jose, California
Anaheim-Santa Ana-Garden Grove, California
Hartford, Connecticut
Houston, Texas
New Orleans, Louisiana
New York, New York-New Jersey
Jersey City, New Jersey

* Model City in group.

Table 4-11

Classification of Small Cities

GROUP 1

Abilene, Texas*
Amarillo, Texas
Odessa, Texas
Midland, Texas

GROUP 2

Asheville, North Carolina
Springfield, Missouri
Springfield, Ohio
Lexington, Kentucky
Sioux Falls, South Dakota
Terre Haute, Indiana
Vineland-Millville-Bridgeton, New Jersey
Sioux City, Iowa-Nebraska
St. Joseph, Missouri*
Roanoke, Virginia
Springfield, Illinois
Bloomington-Normal, Illinois

GROUP 3

Bristol, Connecticut
Reno, Nevada
Norwalk, Connecticut
Columbia, Missouri
Fitchburg-Leominster, Massachusetts
New Britain, Connecticut
Meriden, Connecticut
Lincoln, Nebraska*
Pittsfield, Massachusetts
Rochester, Minnesota
Fargo-Moorhead, Minnesota-North Dakota
Galveston-Texas City, Texas
Champaign-Urbana-Rantoul, Illinois
Gainesville, Florida

GROUP 4

Lynchburg, Virginia
Wilmington, North Carolina
Texarkana, Texas-Texarkana, Arkansas
Tyler, Texas
Monroe, Louisiana
Spartanburg, South Carolina*
Tuscaloosa, Alabama
Owensboro, Kentucky

GROUP 5

Anderson, Indiana
Lafayette, Louisiana
Brockton, Massachusetts
Racine, Wisconsin
Kenosha, Wisconsin
Dubuque, Iowa
Williamsport, Pennsylvania
Santa Cruz, California
Ogden, Utah
Provo-Orem, Utah
Fort Myers, Florida
Gastonia, North Carolina
LaCrosse, Wisconsin
Elmira, New York

GROUP 5 (continued)

Bay City, Michigan
Lima, Ohio
Sherman-Denison, Texas
Sarasota, Florida*
Lawton, Oklahoma
Mansfield, Ohio
Bryan-College Station, Texas
Gadsden, Alabama
Lake Charles, Louisiana
Fort Smith, Arkansas-Oklahoma
Daytona Beach, Florida
Biloxi-Gulfport, Mississippi
Jackson, Michigan
San Angelo, Texas
Waco, Texas
Wichita Falls, Texas
Pueblo, Colorado
Tallahassee, Florida
Topeka, Kansas
Waterloo-Cedar Falls, Iowa
Lafayette-West Lafayette, Indiana
Decatur, Illinois
Billings, Montana
Great Falls, Montana
Green Bay, Wisconsin
Cedar Rapids, Iowa
Boise City, Idaho
Killeen-Temple, Texas
Salem, Oregon
Battle Creek, Michigan

GROUP 6

Modesto, California
Muskegon-Muskegon Heights, Michigan
Muncie, Indiana
Steubenville-Weirton, Ohio-West Virginia
Brownsville-Harlingen-San Benito, Texas
Florence, Alabama
Yakima, Washington
Parkersburg-Marietta, West Virginia-Ohio*
Altoona, Pennsylvania

GROUP 7

Lewiston-Auburn, Maine*
New Bedford, Massachusetts
Fall River, Massachusetts-Rhode Island

GROUP 8

Petersburg-Colonial Heights-Hopewell, Virginia
Durham, North Carolina
Albany, Georgia*

UNCLASSIFIED

Lubbock, Texas
Richland-Kennewick, Washington
Manchester, New Hampshire
Portland, Maine
Danbury, Connecticut
Laredo, Texas
McAllen-Pharr-Edinburg, Texas
Wheeling, West Virginia-Ohio
Atlantic City, New Jersey
Pine Bluff, Arkansas
Savannah, Georgia
Alexandria, Louisiana
Nashua, New Hampshire

Table [illegible]

Classification of Medium Cities

GROUP 1

Albuquerque, New Mexico
Lorain-Elyria, Ohio
Wichita, Kansas
Tulsa, Oklahoma
Hamilton-Middletown, Ohio
Des Moines, Iowa
Davenport-Rock Island-Moline, Iowa-Illinois
Utica-Rome, New York
Greenville, South Carolina
Stockton, California
Orlando, Florida
Charlotte, North Carolina
Little Rock-North Little Rock, Arkansas*
Knoxville, Tennessee
Duluth-Superior, Minnesota-Wisconsin
Peoria, Illinois
Spokane, Washington
Evansville, Indiana
Erie, Pennsylvania
Chattanooga, Tennessee-Georgia
Lakeland-Winter Haven, Florida
Huntington-Ashland, West Virginia-Kentucky-Ohio
Charleston, West Virginia

GROUP 2

Bakersfield, California
Saginaw, Michigan
Flint, Michigan
Baton Rouge, Louisiana
Pensacola, Florida
Colorado Springs, Colorado
Lowell, Massachusetts-New Hampshire
Appleton-Oshkosh, Wisconsin
South Bend, Indiana
Raleigh, North Carolina
Fort Wayne, Indiana
Kalamazoo-Portage, Michigan
Tucson, Arizona
Rockford, Illinois
Vallejo-Fairfield-Napa, California
Eugene-Springfield, Oregon
Bridgeport, Connecticut
Waterbury, Connecticut
Fresno, California
New London-Norwich, Connecticut-Rhode Island
Stamford, Connecticut
York, Pennsylvania
Poughkeepsie, New York
Tacoma, Washington*
Salinas-Seaside-Monterey, California
Wilmington, Delaware-New Jersey-Maryland
Binghamton, New York-Pennsylvania
Santa Barbara-Santa Maria-Lompoc, California
Ann Arbor, Michigan

GROUP 3

Harrisburg, Pennsylvania*
Lancaster, Pennsylvania
Trenton, New Jersey
Johnstown, Pennsylvania
Reading, Pennsylvania

GROUP 4

West Palm Beach-Boca Raton, Florida*
Santa Rosa, California
Canton, Ohio

GROUP 5

Lawrence-Haverhill, Massachusetts-New Hampshire
Worcester, Massachusetts*
Madison, Wisconsin
New Haven-West Haven, Connecticut

GROUP 6

Corpus Christi, Texas
El Paso, Texas
Macon, Georgia
Montgomery, Alabama
Columbus, Georgia-Alabama
Jackson, Mississippi
Newport News-Hampton, Virginia*

GROUP 7

Oxnard-Simi Valley-Ventura, California
Melbourne-Titusville-Cocoa, Florida*
Las Vegas, Nevada

GROUP 8

Charleston, South Carolina
Fayetteville, North Carolina
Augusta, Georgia-South Carolina*

UNCLASSIFIED

Lansing-East Lansing, Michigan
Austin, Texas
Mobile, Alabama
Shreveport, Louisiana
Columbia, South Carolina
Wilkes Barre-Hazleton, Pennsylvania
Scranton, Pennsylvania
Long Branch-Asbury Park, New Jersey
Beaumont-Port Arthur-Orange, Texas

*Model City in group.

It is interesting to look at the outliers in these classifications and to understand what they mean. There were 17 outlier cities among the large cities; all show stronger effects (higher factor scores) than the other cities. A careful look, though, shows several patterns of interest:

> Nashville looks like both groups 5 and 6, highly positive on factors 3, 4 and 5.
>
> Hartford fits group 1, negative on factor 2 and positive on factor 5, but stronger.
>
> Minneapolis-St. Paul, Washington, D.C., and Chicago are all positive on factor 4, negative on factor 5, and form a group which is a stronger version of group 3, particularly the Atlanta-Baltimore subgroup.
>
> New York and Jersey City also form a group--both significantly low on auto use compared to any other cities. Boston and New Orleans are similar, but weaker.
>
> Houston and Buffalo have very low values on VMTP and are highly negative on factor 5. Since VMTP was estimated rather than measured, this could be an error. If it was, Houston would fit group 4 and Buffalo, group 3.
>
> New Brunswick, NJ combined characteristics of groups 1 and 5, negative on factor 2, and positive on factors 4 and 5.
>
> Finally, San Antonio is like no other city: highly positive on factor 3 and negative on factor 5.

This is fairly encouraging: 6 of the outliers fitting groups or the interfaces between groups; 8 forming new groups of their own, which makes them very separate from the main body but similar to each other; and only 3 (Houston, Buffalo, and San Antonio) which do not seem to fit at all. Some of the patterns of the new groups show up in the earlier analysis, but the effects were weaker and the cities were absorbed into group 3.

The results of this classification may be used for a variety of purposes. The modal cities suggest natural case study sites, or locations for demonstration projects; the results of these may then be generalised to other cities in their groups. Further considerations may lead to other choices, data availability represents a case in point. The results of these studies may also be generalised to a larger set of cities. In addition, factor scores may be used to compare cities along the urban transportation related dimensions defined by the factors. Large cities case studies, for example, should be located in Providence, Rhode Island, and Louisville, Kentucky, with the Providence results being applicable to Akron, Ohip; Rochester, Pittsburgh, and so on, and the Louisville results being relevant for Atlanta, Baltimore, Omaha, and so on. Medium city case studies ideally should be located in Little Rock and in Tacoma. Should study results be available for "outliers" cities such as Nashville, Hartford, or San Antonio, the factor scores for these cities should be examined individually. They may indicate that the SMSA--San Antonio, for example--is like no other city, in this case the results cannot be generalized. For the outliers, similarities may be discovered at least along some of the axes, for example, Nashville resembles groups 5 and 6 (of the large city groupings)--but does not fall into them because of an extreme value on Factor 1. Thus, the results have at least limited relevance to Toledo, Norfolk, and other cities in these groups.

APPENDIX A

FACTOR ANALYSIS

Factor analysis is a technique for transforming large bodies of data to a more useful form. Basically, the technique provides a description of the original data set in fewer, and perhaps more meaningful, dimensions. The result can provide a basis for accepting or rejecting certain hypotheses and a simplified description of the data sample. City classification is only one of many applications for factor analysis. In this and other examples of classification, description of the data set within a less complex vector space provides a workable framework in which observations can be grouped together.

Factor analysis was originally developed for use in grouping psychological traits and the greatest number of applications have been in the field of psychometrics. Margaret Hagood first applied the method to regionalization, using data on population and agriculture in the United States.* In the last decade, Brian Berry has been the primary developer of regionalization by factor analysis. In the *Atlas of Economic Development*,** Berry offers one of the earlier and more complete discussions of the value of factor analysis in regionalization for the identification, summation, and condensation of a larger series of detailed information into a relatively smaller group of components or factors. In this exercise, Berry ran a factor analysis on 43 indices for 95 countries. Isolated were four basic patterns characterized by Berry as (1) technological, (2) demographic, (3) income and external relations, and (4) large or small countries. Berry maintained at that time that the "basic similarities were such that four maps would, for certain purposes, do about the same job as 43!"***

A long and detailed series of articles concerning the process of regionalization through the use of factor analysis have followed during the past

---

*Margaret Jarman Hagood, "Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population," *Social Forces*, XXI (March 1943): 287-[illegible].

**Brian J.L. Berry, "Basic Patterns of Economic Development," *Atlas of Economic Development*, Norton Ginsburg, ed. (Chicago: University of Chicago Press, 1961): 110-119.

***Ibid., p. 110.

10 years.* At first, factor analysis was thought of as a type of panacea for the process of regionalization. Later many research scholars have looked upon it as being unapplicable for use in the process of grouping and regionalization.** It would appear that neither of these reactions to the mathematical technique are correct. Factor analysis is a highly subjective method in the sense that the analyst is required to make a number of assumptions about the universe from which the data sample is taken and to choose from extensive alternatives in designing the research. The outcome of the factor analysis, then, is highly dependent on these decisions and should be considered carefully in light of previous theoretical work and research hypotheses. If no hypothesis is formulated and the research design decisions are largely arbitrary, the factor analysis solution is of little value. By the same token, if a proper hypothetical groundwork is laid and the research design is consistent with the hypothesis, the factor analysis technique can be very powerful, since it allows the application of either prior knowledge or hypotheses to be tested. Herein lies the truth about factor analysis: the technique is only as valid as the logic to which it is applied.

THE BASIS OF FACTOR ANALYSIS

As stated previously, factor analysis simplifies the description of a data base by reducing the number of dimensions involved. This transformation is made possible by moderate or high correlation between variables which compose the original data base. When two or more variables are very highly correlated, a single "factor" may describe their variation adequately. In geometric terms, correlation between the original variates indicates non-orthogonal dimensions. In most cases, resolution into orthogonal dimensions will simplify the vector space.†

---

*Brian J.L. Berry, "An Inductive Approach to the Regionalization of Economic Development," *Essays on Geography and Economic Development*, ed. by Norton Ginsburg (Chicago: University of Chicago, Department of Geography, Research Paper No. 62, 1960): 78-107; Brian J.L. Berry, "A Method for Deriving Multi-factor Uniform Regions," *Przeglad Geograficzny*, XXXIII, No. 2 (1961): 263-279; Brian J.L. Berry, Grouping and Regionalizing: An Approach to the Problem Using Multivariate Analysis," in *Quantitative Geography, Part I: Economic and Cultural Topics*, ed. by W.L. Garrison and D.F. Marble (Chicago: Northwestern University Studies in Geography, No. 13, 1967): 219-251. For a more complete bibliography on factor analysis and regionalization, see R.J. Johnson, "Grouping and Regionalizing: Some Methodological and Technical Observations," *Economic Geography* (June supplement, 1970): 293-305.

**An excellent overview and bibliography of the use of Factor Analysis may be found in Philip H. Rees, "Factorial Ecology: An Extended Definition, Survey and Critique of the Field," in *Comparative Factorial Ecology* Supplement to *Economic Geography*, Vol. 47, No. 2 (Supplement), June 1971, pp. 220-234.

†Note, however, that the final (preferred) solution may include non-orthogonal factors.

To illustrate the mathematical framework in which factor analysis is performed, let us consider a data set including N observations in terms of n variables. The data set consists of observations

$$Z_{ij} \qquad j = 1,n; \quad i = 1,N$$

where Z is a normalized variate (mean=0; standard deviation=1)
j denotes the particular variable
i denotes the particular observation.

Relationships between the n variables are measured by correlation coefficients, r:

$$r_{jk} = \sum_{i=1}^{N} Z_{ji} Z_{ki} / N \qquad j = 1,n; \; k = 1,N$$

The transformation from the original dimensions (input variables) to the final dimensions (factors) is expressed as a vector equation:

$$Z_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} \ldots + a_{jm}F_{mi} + a_j U_{ji}$$

where F_j is one of m common factors (i.e., related to two or more input variables)
U_j is a specific factor (related only to the variable j)
a_{mk} is the factor loading (weight) attached to factor k for the variable j.

Given (1) the correlation coefficients (r_{jk}) as a function of input variable values (Z_{ji}), and (2) input variables as a function of the factors (F), the following expression is derived

$$r'_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \ldots + a_{jm}a_{km} \qquad (j \neq k; \; j,k=1,n)$$

where r' denotes the reproduced correlation coefficient.

The factor loadings are derived by equating reproduced correlation coefficients to actual correlation coefficients. The transformation between the input variables and factors is defined at this point.

The solution of the equations described above, however, may or may not be unique depending on the number of restrictive assumptions concerning the correlation matrix. Thinking in geometric terms, again, the factor axes derived may be rotated to various positions which are equally consistent with the correlation matrix. When the initial solution is not unique, it is advantageous to perform such a rotation to obtain the best solution--the preferred solution being chosen according to the criteria of the particular research project.

USES OF FACTOR ANALYSIS

The basic function of factor analysis is to provide a tool with which the fundamental dimensions or interrelationships of a set of variates can be ascertained and measured. As such, the technique is potentially applicable to any field. The applications for such a tool are numerous, but fall mainly into two categories: (1) discerning patterns of interrelationships, and (2) transforming data.

Patterns of interrelationship can be examined with or without prior knowledge of the system sampled. Given prior knowledge, the process is referred to as hypothesis testing (the hypothesis normally concerns the existence of one or more factors). In other cases, factor analysis is applied in an exploratory manner to discover the structure of the data or major sources of variation. As an example of exploratory applications, consider the first study ever performed on crime. The application of factor analysis to the data base including economic, demographic, land use, crime and enforcement data would prove very useful--not in establishing causes but in identifying relationships to be further investigated--identifying a sphere of influence.

Transforming data by factor analysis requires substantially more knowledge about the population being analyzed. In fact, such an application presumes a hypothesis test within the factor analysis if the variable grouping is not known. The objectives in transforming data are numerous. In preparing for a regression analysis, for instance, it is often desirable to reduce the total number of variables and/or to resolve colinear variables into orthogonal measures. Another application of factor analysis is in devising a scale by which individuals or groups can be rated. By attaching weights, the individual characteristics are transformed into scores. Scores devised in this manner are often used to classify the data observations or to provide a more concise description of the data base. In data transformation by factor analysis, then, the primary objectives are usually parsimony, orthogonality, and/or interpretability. One of the strengths of the factor analysis technique is in its capacity to accept criteria with respect to each of these objectives.

RESEARCH DESIGN -- ALTERNATIVE TECHNIQUES

Many alternative techniques have been devised in the development of factor analysis. Although each new method has improved on the last in some way, no single method is regarded as better than all others. In fact, none is applicable to all research efforts. The methods differ in basic assumptions about the data base as well as in solution techniques. As a result, the value of any given application of factor analysis rests in the ability of the researcher to choose the method most appropriate to his research. Rummel designates 10 items to be considered in making this choice.[*]

---

[*] R. Rummel, Applied Factor Analysis (Evanston, Ill.: Northwestern University Press, 1970): 324-328.

(1) grouping requirement--some techniques require input variables to be grouped together according to the research hypothesis;

(2) number of factors--some techniques are constrained to a given number of factors (normally one or two) while others deal with the generalized case of multiple factors;

(3) variable complexity--the major option here is between identifying all relevant factors and limiting the solution to a small number of dominant factors for each variable;

(4) factor complexity--refers to the number of variables having at least moderate loadings on a given factor. The degree of complexity reflects the choice of general factors, which load on all variables, grouping factors, which load on more than one but not all variables, and specific factors, each of which is unique to a particular variable;

(5) factor variance contribution--the researcher must decide whether the factors should all contribute equally to explaining variance, the first should explain the most variance with others explaining a decreasing proportion, etcetera;

(6) sensitivity to error--not all methods give specific consideration to measurement and stochastic error and those which do estimate the error use differing approximations;

(7) determinateness--refers to the uniqueness of the solution. This criteria implies a trade-off since certain assumptions are required in order to reach a unique solution;

(8) resources--some methods are not feasible without the aid of high-speed computers;

(9) hypotheses--the method must be consistent with any hypotheses. Normally this is a question of factor and variable complexity;

(10) theory--special case where factor analysis is applied to geometry.

These 10 criteria are applied in the research design process to choose an approach and solution technique from a large number of alternatives. Discussion here will center on the inherent assumptions of three alternative approaches. Alternative solution techniques are applied within the three approaches but will not be enumerated here. The reader is referred to Rummel* for a concise discussion of these methods.

---

*Rummel, op.cit., pp. 333-347, 101-313.

The two-factor approach is used where one underlying factor is anticipated. The approach assumes a portion of the variation in each variable is related to a single common factor. Variance not explained by this common factor is called unique variance and is attributed to individual specific factors for each variable. This approach constrains the factor loading to be similar to that illustrated in Figure A-1 where each variable loads on the common factor and one specific factor. This method provides a unique solution, given the assumed factor structure. At first glance, the method appears very restrictive since it only deals with a single nonspecific factor. In practice, however, the method can be applied to each of several groups of variables, thus resulting in multiple factors. Scaling and/or orthogonal factors could then be accomplished by application of a multifactor approach. The advantage in this procedure over an ordinary application of a multifactor approach is strict adherence to the research hypothesis.

Figure A-1

Two-Factor Approach*

		F_1	F_2	F_3	F_4	F_5	F_6
Variables	1	x	x				
	2	x		x			
	3	x			x		
	4	x				x	
	5	x					x

*Crosses indicate moderate to high loadings.

The bifactor approach is similar to the two-factor framework in that each variable is described in terms of a common factor and one other factor. The second factor, here, however, is a group factor rather than a specific factor. The bifactor solution is thus similar to the loading shown by Figure A-2. The resulting factors are mutually orthogonal. This approach also provides a unique solution but requires a hypothesis concerning variable interrelationship prior to the factoring process. It is also constrained to disallow overlapping variable groups.

Figure A-2

Bifactor Approach[a]

Variables	F_1	F_2	F_3	F_4
1	x	x		
2	x	x		
3	x		x	
4	x		x	
5	x		x	
6	x			x
7	x			x

[a] Crosses indicate moderate to high loadings

The multifactor approach is a generalization of the two-factor and bifactor theories. Because it does not require *a priori* assumptions about the number and structure of factors, the approach can be applied to any data set with little or no knowledge of underlying factors (abuses resulting from this freedom will be discussed later). Most of the multifactor methods do not derive a unique solution, therefore, these methods involve two steps: (1) extraction of initial factors, and (2) a search to find the optimal result. The search involves transformations of the initial solution and is referred to as a rotation to a terminal solution.

Methods for extracting initial factor solutions fall into two categories: principal component analysis and classical factor analysis. The greatest difference between the two is that classical factor analysis attempts to estimate data error whereas principal component analysis utilizes exact mathematical transformations of the data.

Classical factor analysis presumes both common and unique determinants in each variable. The unique portion of the variance includes all sources of error and must be estimated based on assumptions concerning the number of factors involved. In practice, the estimated quantity is communality, that portion of the variance attributable to common factors. Because the factors are derived from estimated communalities, the solution comprises inferred factors.

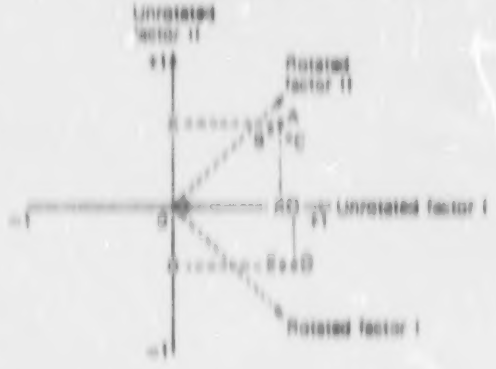
Principal component analysis, on the other hand, yields defined factors--exact transformations of the input data. Errors in the data are not specifically considered and, therefore, need not be estimated. The general structure of factors need not be assumed for this type of analysis. The factor solution by this method consists of as many factors as the original number of variables. Factors are chosen so that the first factor explains the largest proportion of variance, the second explains the largest proportion of the remaining variance, and so on. All of these factors do not appear in the final solution, however, since a lesser number will explain most of the variance. The researcher must consider how many of the factors are to be retained for rotation to a final solution.

Given an initial solution of either defined factors or inferred factors, the researcher normally proceeds with a search to find the most desirable solution in view of the particular research criteria. In geometric terms, the original data observations represent points in space and the initial factor solution is a set of coordinate axes devised to describe these points. Although the number of factors in the initial solution is chosen according to the number of orthogonal dimensions in the vector space, the orientation of the factor axes is largely arbitrary. By rotating the factor axes, it is often possible to obtain new solutions with fewer factors, more meaningful factors, or both, which describe the data base equally well.

Figure A-3 shows a simple example of factor rotation. Note the variables are unchanged--the factor axes rotate. The result of the rotation is to simplify description of the variables. Whereas before the rotation every variable loaded heavily on both factors, the rotated factors clearly show one group loading mostly on the first factor and another which loads heavily on the second factor. The rotated factors are easily interpreted since each factor is very highly correlated with a known set of variables. In this example, the factor axes are rotated together. It is also possible to rotate the axes with respect to each other by an oblique rotation. Figure A-4 demonstrates an oblique rotation of the axes in Figure A-3. The number of factors remains the same, however the factor loadings are concentrated still further. In addition, the oblique rotation provides information on factor correlations.

Figure A-3

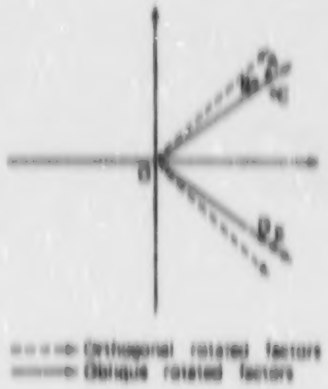
Orthogonal Factor Rotation[a]



[a] Source: Nie, Bent & Hull, Statistical Package for the Social Sciences, New York: McGraw Hill Book Company, 1970, p. 222.

Figure A-4

Oblique Factor Rotation*



*Source: Nie, Bent & Hull, Statistical Package for the Social Sciences, New York: McGraw Hill Book Company, 1970, p. 223.

In dealing with more than two dimensions the choice of a final solution is far more complex than is indicated by the example just related. Where the vector space comprises many dimensions, the objectives of parsimony, interpretability and orthogonality are often in opposition. In many cases, more meaningful factors can be attained by increasing the number of factors and/or performing oblique rotations. Interpretability can be sacrificed to achieve parsimony as well. Factor rotation is a very subjective process due to this trade-off between statistical simplicity and conceptual meaningfulness. Furthermore, if meaningful factors is a criterion, selection of a final solution is not based entirely on geometry. The analyst makes a subjective decision, influenced by prior knowledge of variable interrelationships, research goals, and personal biases

ASSUMPTIONS, LIMITATIONS, AND ABUSES

Factor analysis has weaknesses, as do most quantitative methods, but most of the criticisms launched against the technique results from an incomplete understanding of the methods and their proper use.

Two aspects of factor analysis constitute real shortcomings and should be mentioned here. First, there is an inherent bias in factor analysis solutions resulting from the use of normalized scores as inputs to the analysis. The scores are normalized to alleviate potential bias toward large numbers and large factors. In some cases, however, the result may be to attach excessive weight to variates with small variances relative to their magnitudes or underweight variates with relatively large variances.

The second deficiency pertains to classical factor analysis and its lack of a precise solution. The common factor solution requires that elements on the main diagonal of the correlation matrix be replaced by estimates of communality prior to the extraction of initial factors (recall communality is that portion of the variance which is related to other variables). Although many arbitrary procedures have been devised to estimate communality, the true values are unknown; all common factor analysis solutions are imprecise in this respect.

Aside from these two shortcomings, misunderstanding and, therefore, frequent misuse has been responsible for most present skepticism regarding the use of factor analysis.

The most widespread criticism of factor analysis has perhaps been that the technique is largely arbitrary--that different analysts may arrive at different solutions even though they might use the same data and techniques. In previous applications, such criticism may be well-founded, although in theory the technique is not arbitrary. Factor analysis is very subjective because it relies on the analyst's judgement in numerous research decisions with respect to inputs to the analysis, techniques of factoring and solution criteria. If the analyst applies previous knowledge of the system being sampled and uses sound logic in developing the research design, the analysis is anything but arbitrary. Furthermore, if the research is properly documented with assumptions and techniques clearly noted, another analyst can repeat or modify the analysis based on the logic of the original research design.

The researcher's alternatives begin with the content of the data base to be analyzed. Three characteristics of this body of data are important. Most obvious is the combination of variables chosen for inclusion. If all of the relevant variables are not included, the factor solution can be misleading. The particular cases cited will influence results as well. Just as in any sampling problem, care must be taken to represent the population as closely as possible with the sample. Choice of the unit of analysis is equally important. For instance, a classification of SMSAs (Standard Metropolitan Statistical Areas) would likely differ from a classification for the same cities by city proper since the relationship between city proper and SMSA is neither constant nor defined. As these determinants of the input data affect the research conclusions, it is essential that the analysis and conclusions include an assessment of the meaning of the data and consequences of variable choice.

The number of alternative assumptions for and methods of extraction of factors are considerable.* The important point to note is that a unique solution is obtained only by making restrictive assumptions. Therefore, to arrive at identical solutions, two analysts must make the same assumptions (i.e., utilize the same technique). A report of the research, then, is not complete without documentation of the logical development of the methodology for factor solution.

---

*See, for instance, Rummel, op.cit.

Decisions made in the final stages of the analysis are better recognized with respect to their effect on the outcome of the research. At this point, the researcher determines both the scope of the search for a preferred solution (rotation technique) and the criteria by which the preferred solution is established.

The formulation of research hypotheses is essential to a consistent and informed research design since it is the hypothesis which expresses both the research goals and prior knowledge of the system. In too many instances, factor analyses have been performed with no hypotheses--the result being that the data is used to derive a theory rather than to test it. Such abuses of the technique of factor analysis cannot hope to produce valid results since the choice of input data generate meaningless and even dangerous conclusions.*

Lack of a prior hypothesis for factor analysis means insufficient criteria for evaluation of alternative factor solutions. Having used the data base to develop a theory, there is no check on the theory remaining. Armstrong suggests, therefore, that the proper use of the factor analysis technique should include prior assumptions concerning

- relationships among the variables
- number of factors
- types of factors
- variables to be included in data base
- relationships between the resulting factors
- meaningful communality estimates for the problem.+

---

*Armstrong provides a fictitious example of a young analyst's attempt to investigate the nature of a sample of metal blocks. Using a data base of easily acquired measurements, he found the important characteristics of the metal to be intensity, shortness and compactness.** Actual applications of the technique can be almost as amusing. Cattell "... attempted to discover the primary dimensions of culture by obtaining data on 72 variables for each of 69 national cultures. The 12 basic factors which were obtained...included...cultural assertion, enlightened affluence, thoughtful industriousness, bourgeois philistinism, and cultural disintegration."*** It is not likely Cattell hypothesized these factors. In fact, had either of the examples just related included a proper hypothetical framework, interpretation of the factor solution might have resulted in different conclusions.

**J. Scott Armstrong, "Derivation of a Theory by Means of Factor Analysis or Tom Swift and His Electric Factor Analysis Machine," The American Statistician (December 1967): 17-21.

***R.B. Cattell, "The Dimensions of Culture Patterns by Factorization of National Characteristics," Journal of Abnormal and Social Psychology 44 (1949) pp. 443-469. As cited by Armstrong.

+Armstrong, op.cit.

APPENDIX B

SURVEY OF CITY AND COMMUNITY CLASSIFICATIONS

Although considerable resources have been directed toward developing community typologies, few of the resulting classifications have been applied to further research or practical problems. One reason is that every potential use of a community typology has specific requirements in terms of city characteristics considered, and the universe of communities to be investigated. No classification, then, is useful in every case. In addition, there has been a wide range of methods applied to city classification with respect to the choice of relevant city characteristics, the measurement of those attributes, and grouping cities in a multidimensional space. This diversity in methodologies reflects a certain amount of dissatisfaction with the methods used and results obtained to this point.

With respect to the present research, none of the studies reviewed provided an appropriate typology of cities, however the literature demonstrates the variety of classificatory systems in terms of (1) characteristics by which the communities are classified, (2) variates utilized to measure these attributes, and (3) criteria for grouping similar observations.

A comprehensive description of any community would require an infinite number of dimensions. Of necessity, every classificatory system must rely on an approximate description based on a workable number of dimensions. If the objectives of the classification are very specific, the data base may also be specific. However, as more has been learned about the complex interrelationships of city attributes, classificatory studies have been progressively broader in scope. The studies investigated here have distinct orientations which have been categorized as follows: (1) economic function, (2) hierarchical structure, (3) growth potential, (4) social factors, and (5) land use, environmental quality, and quality of life.

Economic function refers to the city's primary area of contribution to the national economy in terms of production, sales, and service. Assessment of the function of cities has been primarily through employment data, though various transformations of raw data have been utilized (see Bahl, Firestine, 1971, for tests of alternative measures). Douglas (1967) included payroll, value added, and capital expenditure measures in his analysis, which resulted in a descriptive summary of the cities' industrial characteristics rather than a typology. As an example of functional city types, Nelson (1955) distinguished the following nine groups: (1) manufacturing, (2) retail trade, (3) professional service, (4) transportation and communication, (5) personal service, (6) public administration, (7) wholesale trade, (8) finance, insurance and real estate, and (9) mining. Two types of functional classifications have been performed. Most of these typologies are developed according to the field in which the city shows the greatest activity concentration. In some cases, such as Alexandersson (1956), the cities are classified separately for each industry. Such multiple classifications tell little about the overall composition of the city.

Hierarchical classifications have recently become popular, being oriented toward the emerging urban system of communities. Descriptive analyses of spatial structure, land use, and hierarchical structure have been presented by Lampard (1968) and Sacks, et al (1972), but Berry (1973) is the first to derive a hierarchical structure based on "economic power." Factor analysis provided the measure of economic power, which was used to identify five levels of the hierarchy:

1. metropolis
2. major wholesale-retail center
3. complete shopping goods center
4. partial shopping goods center
5. complete convenience goods center.

Since the largest cities are now quite diversified, the hierarchical classification seems more useful than the functional system.

Very much related to hierarchical structure is growth potential and spatial structure, one aspect of urban areas singled out for classification. Growth potential is determined by many factors and is thus very difficult to quantify. Typological studies may be of great value in this area of research. Ray (1969) has applied factor analysis techniques to variates representing economic and cultural characteristics in Canada and has discerned three spatial patterns: "a national heartland-hinterland pattern, an urban hierarchy pattern, and an intermetropolitan axis pattern." Dutton (1970) actually estimates growth potential for the U.S. Using income and population potentials, Dutton maps growth potential contours for the entire country. Although these studies do not generate city classifications as such, they do illustrate basic patterns to be considered in city classification.

The emphasis in research discussed above is on economic function, power, and potential. Equally important, however, to the character of a community is the population. Although many studies have ignored social and demographic characteristics, Berry (1972) has concluded from a comprehensive review of the literature that the basic characteristics of a city are economic function and sociocultural factors. Sociocultural factors identified include functional size, socioeconomic status, stage in life cycle, nonwhite and foreign-born populations, recent growth in population or employment, and composition of the labor force. Underlying these factors are the basic dimensions of social status and age structure, or stage in life cycle. A study by Hadden and Borgatta (1965) provided the basis for many classifications by sociocultural factors. "Hadden and Borgatta were very critical of the economic function approach to city classification. They deliberately include variables which would produce some factors representing types of economic specialization, and found that very little of the variance of the other conditions included in the study accrued to these specialization factors."* Their analysis identified three primary factors: socioeconomic status level, nonwhite population, and age composition. Other works emphasizing social and demographic factors include Nelson (1957), Meyer (1972), Harries (1973), and Hughes (1973).

Recently, there has been great interest in environmental quality and quality of life and three classificatory studies may provide the basis for further research in this area. Coughlin (1970) performed a factor analysis for 101 metropolitan areas on 60 indicators of environmental quality and quality of life. The analysis was biased toward social and economic characteristics since data on physical conditions was sparse. The primary axes identified were called:

1. cosmopolitan and affluent goal axis
2. middle America axis
3. sound economic structure vs. open space and lack of hypertension axis
4. clean air axis
5. health goal axis.

Jonassen and Sherwood (1960) performed a similar study for 88 counties of Ohio without measures of physical environmental quality. Seven factors were derived from the data base of 82 variables: (1) urbanism, (2) welfare, (3) influx, (4) poverty, (5) magni-complexity, (6) educational effort, and (7) proletarianism. More relevant is Berry's study (1974), Land Use, Urban Form, and Environmental Quality, which provides a classification based on social, economic and environmental characteristics. Although there are significant weaknesses in both the data base and the methodology used for this research, Berry has made a beginning and provides inspiration, as well as a core data base, for future research.

Classification of communities requires the measurement of relevant attributes. In the studies reviewed, such measurement was performed at very different levels of sophistication. Alexandersson (1956) used raw data on chosen measures of economic activity. Many authors have combined measurable variables using either arbitrary or theoretically derived transformations (see Bahl, Firestine, Phares: 1971; Shevky: 1955), and studies using larger bodies of data have employed factor analysis to scale off the basic dimensions. Duncan and Reiss (1956) performed a qualitative factor analysis on data prepared for the 1950 Census of Population. Four underlying causal factors were assumed, rather than derived: size of community, spatial organization of communities, community growth and decline, and functional specialization of communities. Each of these factors was explored separately in terms of the demographic measures available. Quantitative methods of factor analysis are useful in identifying basic dimensions from the data base, and these are the techniques most often utilized in comprehensive city classification. Applications of factor analysis in this field have achieved a high level of sophistication (see, for instance, King & Jeffrey: 1972).

---

*David M. Smith, "The Classification of Cities," The Geography of Social Well-Being (New York: McGraw-Hill, 1973): pp. 24-28.

Given a coordinate description of communities to be classified, whether along dimensions included in the data base or derived from the data base, some means of grouping the observations must yet be applied to obtain a typology of cities. Several different methods of grouping have been utilized in the past. The simplest is to develop separate classifications along each dimension, as in Kneedler (1945). A set of graphs plotting observations against two dimensions at a time provides a basis for one classificatory system across all dimensions, although this method requires a great deal of judgement (see Moser and Scott: 1961). A third method requires a tree of classes and subclasses, each level of which corresponds to one dimension (see Forstall: 1970). This type of classification is easy to perform but requires setting of priorities between city attributes. Other more objective methods are also used to cluster similar observations. Elementary cluster analysis entails the grouping of observations according to their correlation coefficients (Armen: 1972; Kass: 1973; Marshall: 1975). Another method devised by Jones (1968), uses rotation of the factors to achieve maximum modality. Arbitrary cut-offs along the bimodal axes are used to define clusters (Jones and Jones: 1970). The Ward algorithm does not require arbitrary cut-offs. Cities are clustered according to the following criterion: for any number of groups, the cities within groups are maximally similar with respect to within group variation (Ward: 1963; Bruce and Witt: 1971; King: 1966). A refinement of the Ward algorithm has been instituted by Sutton, et al (1974).

Of the research efforts discussed above, none have produced typologies appropriate to the present research in terms of the universe of cities classified, city characteristics emphasized, and classification methodology. Those which approach the needs of the present research most closely are Bruce and Witt (1971), Sutton et al (1974), Coughlin (1970), and Berry, et al (1974). Bruce and Witt, and Sutton et al have classified cities rather than urban areas and neither of these includes environmental attributes in the analysis. Coughlin, on the other hand, has classified cities according to quality of life measures only. Berry's analysis includes socioeconomic, functional, land use and environmental attributes and classifies 76 urban regions (SMSAs).

Alexandersson, Gunnar. *The Industrial Structure of American Cities: A Geographic Study of Urban Economy in the United States*. (Lincoln: University of Nebraska Press, 1956).

This study is actually not one classification, but several. The methodology entails a simple aggregation of data and mapping rather than factor analysis and cluster analysis. The cities are described by census data on industry and employment. Classifications are performed separately in each industry described by the census according to employment figures. A classification according to percentage of people employed in manufacturing overall is also presented.

Armen, Garbis. "A Classification of Cities and City Regions in England and Wales, 1966." *Regional Studies* (Vol. 6, No. 2, June 1972): pp. 149-182.

"An outline of the development of the concepts of the city and the city region is followed by the physical delimitation and classification of 100 examples of each, on the basis of some 130 characteristics relating to people, place, and activity as in 1966. In contrast to factor analysis, the illustrated method of classification--elementary cluster analysis--lays equal emphasis on all selected characteristics and accepts the occurrence of overlap between classes relating to complex entities like cities and city regions. Each classification results in a hierarchy of three nesting groups: "divisions" (e.g. service and manufacturing cities and city regions) which comprise "classes" (e.g. ports, holiday resorts, etc.) which are made up of "subclasses" (e.g., large ports, medium ports, etcetera). The broad similarity of results for cities and city regions suggests that the character of a city region is at present largely determined by the characteristics of the city upon which it is centered."

Bahl, Roy W., Robert Firestine and Donald Phares, "Industrial Diversity in Urban Areas: Alternative Measures and Intermetropolitan Comparisons." *Economic Geography* (Vol. 47, No. 3, July 1971): pp. 414-25.

The study compares four alternative methods of measuring industrial diversity. An empirical analysis including the four methods used all 212 SMSAs defined at the time of the 1960 Census. Two-digit, 41 industry SIC taxonomy was used. Each of the methods used some form of deviation of employment ratios summed over all industries. Indices were computed for each of the 212 SMSAs. Conclusions: deviations from an estimated minimum requirement provided the best results (as opposed to national average, etc.).

Berry, Brian J.L., *Growth Centers in the American Urban System.* (two volumes) (Cambridge, Mass.: Ballinger Publishing Company, 1973).

The study provides a classification of U.S. cities on the basis of importance as a market center. "All urban places in the United States with populations exceeding 10,000 in both 1960 and 1970 were subjected to a 97-variable multiple factor analysis, the first factor of which indexed the relative 'economic power' or 'hierarchical status' of each center. This analysis is described in the *City Classification Handbook*."[*]

On the basis of these index values and, for smaller places, population size, all of the nation's urban areas with populations exceeding 1,500 in 1960 were stratified into a hierarchy with the following levels:

metropolis
major wholesale-retail center
complete shopping goods center
partial shopping goods center
complete convenience goods center

Hierarchical breaks were based largely on comparison of factor scores for the Upper Great Lakes Region with a hierarchical grouping developed by Borchert and Adams[**] and several other urban area classifications.

---

[*] "Latent Structure of the American Urban System, with International Comparisons." pp. 11-60.

[**] John R. Borchert and Russell B. Adams, *Trade Centers and Tributary Areas of the Upper Midwest.* Upper Midwest Economic Study Urban Report No. 3 (1963).

Berry, Brian, J.L. et al. Land Use, Urban Form and Environmental Quality (Chicago: Department of Geography, University of Chicago, 1974).

The purpose of this study was to isolate key relationships between urban form, land use and environmental pollution. As the research design called for the study of detailed local relationships based on a sample of representative urban regions, a pollution-sensitive typology was developed. Principal component analysis was applied to a sample correlation matrix of urban regions rather than the correlations of variables describing the urban regions. The resulting factors, therefore, identify groups of cities rather than common characteristics of cities and interpretation of these groups is difficult. In order to describe these groups by their characteristics it is necessary to trace the common characteristics of the cities within each group through the original data set. Detailed tables provided an insight into the basic characteristics which differentiated the groups, however, quantification of their relative importance was not achieved.

Berry, Brian J.L. "Latent Structure of the American Urban System, with International Comparisons," City Classification Handbook

Berry provides a simple explanation of the theory of factor analysis as it applies to city classification. This is followed by an excellent review of underlying factors determined by previous studies. Discussion of the individual factors relates to two basic types of characteristics--system-wide sociocultural factors and economic function. Sociocultural factors include functional size, socioeconomic status, stage in life cycle, non-white and foreign-born populations, recent growth in population or employment, female participation in the labor force, and elderly males in the labor force. Functional town types include manufacturing towns, mining towns, college towns, military installations, and service centers. Berry concludes from a number of analyses that the conventional functional classification of cities is becoming progressively less useful as the number of metropolitan centers grows, since these centers are multifunctional. Generality of the underlying factors is demonstrated by describing several independent factor analyses of United States cities and foreign urban systems as well. The following conclusions were reached:

1. "The economic base of urban centers tends to act independently of other urban structural features..."

2. "Every urban system is organized system-wide into a hierarchy of centers based on aggregate economic power..."

3. "In every society the principal dimensions of socioeconomic differentiation are those of social status and age structure, or stage in life cycle. However, only at the highest levels of development do these factors appear to operate independently..."

Bottiny, Walter H. and Beatrice T. Goley. "A Classification of Urbanized Areas for Transportation Analysis." Highway Research Record #194 pp. 32-61.

"As an initial step in the development of an urban area transportation typology, 213 urbanized areas (1960) were tentatively segregated into four population size groups and then subdivided, on the basis of population density, into three density classes for each group. This approach was based on the theory that automobile availability and the mode of transport used in the work trip are functions, primarily of population and density.

Analysis revealed the inadequacy of population density in explaining urban area variations in most of the selected indicators of transportation and socioeconomic characteristics.

An inferential statistical method, the stepwise multiple-regression procedure, was used to analyze the interrelationship between six transportation variables (dependent) and 12 socioeconomic variables (independent) in each of the four population size groups. Several of the equations developed have potential use for estimating certain indicators of automobile availability and travel characteristics.

A regrouping of urbanized areas, using as the criterion a ratio of the number of automobiles available to the number of persons employed, resulted in urbanized area types which in each population size group display distinguishing transportation characteristics as well as characteristics underlying socioeconomic factors and geographic location."

Bruce, Grady D. and Robert E. Witt. "Developing Empirically Derived City Typologies: An Application of Cluster Analysis," Sociological Quarterly (Vol. 12, No. 2, Spring 1971): pp. 238-46.

Previous functional classifications of cities:

uni-dimensional: Harris (1943), Ohlson (1945), Jones, Forstall & Collver (1963).

multi-dimensional: Nelson (1955), Angell (1951), Hagood, et.al. (1941), Alexandersson (1956), Duncan and Reiss (1956), Hadden and Borgatta (1965).

even multi-dimensional classifications have performed categorization along one dimension rather than factors.

The authors accept the Hadden and Borgatta results as useful with certain modifications. Two of eight factors are eliminated following reinterpretation of the results. Then the 83 cities are clustered given the following criterion: "for any given number of groups, the cities within groups are maximally similar with respect to within group variation." The resulting 13 clusters appear to be meaningful based on measures of the six factors.

Reference is made to two other studies which have employed this method of clustering successfully: McLean and Bruce (1969) and Bruce (1969).

Coughlin, Robert E. Goal Attainment Levels in 101 Metropolitan Areas. RSRI Discussion Paper Series No. 41 (Philadelphia: Regional Science Research Institute, 1970).

The quality of life of 101 SMSAs with populations of 250,000 or more is investigated in this research. A factor analysis was performed on 60 indicators measuring the following characteristics:

purity of air
political & government well-being
education
income & wealth
housing integration
normal family structure
availability of services
lack of stress & other environmental related disease
structural soundness of economy
open space
health
housing
economic equality
employment rate
safety
availability of transit
dominance of central core

Data on physical conditions were scarce relative to data on social, economic and health characteristics. Few negative correlations were discovered; no evidence of trade-offs. Income was found to be highly correlated with many indicators --physical characteristics being the exception (e.g. open space). Results of initial factor analysis were adjusted to eliminate the effects of population size.

Douglas, Robert C. Selected Indices of Industrial Characteristics for U.S. SMSAs, 1963 (Philadelphia: Regional Science Research Institute, RSRI Discussion Paper Series No. 20, December 1967).

This study attempts to summarize the economic structure of the 67 largest U.S. SMSAs for 1963 using several indices:

payroll per employee
value added per employee
value added minus payroll per employee
new capital expenditure per employee
diversification

All but the first and last come from the 1963 Census of Manufacturers. Payroll per employee and diversification indices used the same source of data but required theoretical input and mathematical transformations. Construction of the payroll per employee index utilized both local and national payrolls in each industry and employment in each industry to devise a ratio of standardized payroll per employee in area j to actual payroll per employee in area j. The standardized payroll was based on local employment in each industry and national average wage rates in each industry.

The diversification index was based on a group of location coefficients, or concentration indices for the separate industries.

The location coefficient for industry i in SMSA j is simply the percent of j's employment in i divided by the percent of employment in all 67 SMSAs found in industry i.

The diversification index was derived from this group of location coefficients using a Lorenz curve. This method measures the extent to which the location coefficients for a given SMSA differ from a hypothetical group of location coefficients all equal to 1.

Results consist of tabulated descriptions.

Duncan, Otis D. and Albert J. Reiss, Jr. *Social Characteristics of Urban and Rural Communities, 1950* (New York: John Wiley and Sons, 1956).

This study was an important contribution to the systematic study of communities. Although the research develops no new theory and tests no specific hypotheses, it does reveal some of the research possibilities in census data and begins to interpret the data. The study is based primarily on the statistics in the special report of the 1950 Census of Population entitled, "Characteristics by Size of Place." The analysis is based on the hypothesis that there are orderly processes and relationships which produce the visible diversity in community life. Census data are taken to represent these observable characteristics and four underlying causal factors are assumud:

size of community
spatial organization of communities
community growth and decline
functional specialization of communities

Each of these factors is discussed separately in terms of the demographic measures of the census. Findings are extensive and qualitative. The analysis amounts to a qualitative factor analysis.

Dutton, Geoffrey H. *Macroscopic Aspects of Metropolitan Evaluation*. Harvard Papers in Theoretical Geography, Geography of Income Series: Paper No. 1 (Cambridge, Mass.: Harvard University Graduate School of Design, Laboratory for Computer Graphics and Spatial Analysis, 1970).

Dutton divides U.S. into regions which are determined by population and income potentials and are intended to reflect extrinsic desirability (i.e., desirable in terms of fulfilling demands beyond those relating to physical survival: art, education, science, religion, etc.) These are found to correlate very well with surveys performed by Gould to assess subjective spatial preferences. Extrinsic desirability is actually a ratio of income potentials to population potentials, each of which is based on a gravity-field concept.

Forstall, Richard L. "A New Social and Economic Grouping of Cities," *The Municipal Yearbook* (Washington, D.C.: The International City Management Association, 1970): pp. 102-123.

This article presents a classification of 1761 U.S. cities of 10,000 or more population in 1960, based on their social and economic characteristics. This classification has a much broader scope than previous classifications published in the yearbook, which related almost entirely to the city's function in terms of business activities. A multivariate factor analysis is performed here on 97 variables on which data was provided by the 1960 Census of Population and the 1963 Census of Business. Fourteen factors resulted of which 6 were retained for use in classification. The classification of cities was both quantitative and qualitative. Cities were classified according to the following items (in order of priority):

- region
- metropolitan status
- size
- specialization in mining, colleges or military
- socioeconomic status
- state in family cycle
- proportion of nonwhites
- proportion of the labor force in manufacturing

these are not factors
but independent measures

The final classification included 241 groups.

Friedman, Judith J. "Variations in the Level of Central Business District Retail Activity Among Large U.S. Cities: 1954 and 1967." Land Economics (49:3, August 1973): pp. 326-335.

"The effects of relative isolation, time of development and population size upon CBD retail sales are examined for 1954 and 1967 using multiple regression analysis and, to a limited extent, path analysis. Comparing the regression coefficients in parallel equations for the two years suggests the change in the impact of each of these three independent variables during this 13-year period. Various other variables such as office activity are then added to the 1967 equation one by one."

Conclusions: "CBD retail activity is part of a complex of interdependent economic activities. Certain city characteristics promote the development of such a complex: CBD retail sales increase with the city's relative isolation, the length of time since it first became large, and the SMSA population size. Cities which are relatively isolated and those which are more populous have more office activity; such cities and those which developed early also have more such attractions as museums in their CBDs. CBD sales increase with the amount of office activity, and CBDs with other attractions, especially hotels, also have higher sales."

Golob, Thomas F.; Eugene T. Canty; and Richard L. Gustavson. Toward The Stratification and Clustering of Metropolitan Areas. Paper presented at the Operations Research Society of America Meeting in Dallas, May 1971.

"This paper discusses a procedure by which to classify metropolitan areas into relatively homogeneous groups on the basis of a selected criterion, identify the most representative areas within each group and describe the relationships between areas within the same group and within different groups. The procedure involves the systematic application of multivariate statistical tools including canonical correlation, factor analysis, cluster analysis and discriminant analysis to an extensive data base on metropolitan areas."

The proposed technique is applied to readily available measures of social and economic structure to classify metropolitan areas with respect to their transportation needs. The original data set included 76 variables from the following areas:

population size & density
age of the area
concentrations of population and employment in the
  central city
family size and age distributions
auto ownership levels
retail trade activities
service sector activities
transportation system usages
housing type and quality
employment by occupations
income distributions
educational levels
wholesale trade activities
manufacturing activities
land uses

The data set was reduced to 12 variables for analysis due to high intercorrelations and incomplete data on certain measures. All of the above areas were represented in the analysis except land use. One-hundred ninety-eight metropolitan areas were grouped into six clusters.

Hadden, Jeffrey K. and Edgar F. Borgatta. *American Cities: Their Social Characteristics* (Chicago: Rand McNally, 1965).

This research attempts to determine by factor analysis the underlying dimensions of cities, assuming a broad socioeconomic structure. The data base described cities with populations over 250,000 (in 1960) with 65 measures of demographic characteristics, income, education, housing, employment, economic activity, location, and age of the city. Three major factors were identified: socioeconomic status level, non-white population, and age composition. Factor scores were utilised to develop profiles of the cities rather than a classification.

"Hadden and Borgatta were very critical of the economic function approach to city classification. They deliberately included variables which would produce some factors representing types of economic specialisation, and found that very little of the variance of the other conditions included in the study accured to these specialisation factors."*

Harries, Keith D. "Population Change Trends in Large U.S. Metropolitan Areas, 1960-70: A Review." *The Journal of Geography* (72: 9, December 1973): pp. 19-30.

"Seven hypotheses relating to change in metropolitan areas in the decade 1960-70 are developed from literature pertaining to the preceding decade. Analysis indicates that most of the trends apparent in the 1950's have continued in the 1960's, including racial polarization, the effects of age and location of SMSAs, the uniqueness of the South, and rapid growth in the Southeast and West. Several hypotheses were rejected: (1) black suburban populations remained constant, (2) suburbs in the South are losing non-whites, and (3) SMSAs in the South decentralize more slowly than elsewhere. The influences of annexation, migration, and natural population increase on metropolitan change are briefly discussed."

The analysis comprised simple comparisons of % change. The SMSAs considered here are the 56 with a population of half a million or more in 1970.

---

*David M. Smith, "The Classification of Cities," *The Geography of Social Well-Being in the United States* (New York: McGraw-Hill, 1973): pp. 34-38.

Harris, Chauncy D. "A Functional Classification of Cities in the United States." *Geographical Review* (Vol. 33, No. 1, January 1943): pp. 86-99.

Classification of cities is by activity of greatest importance in each city. Data utilized comprises employment statistics for 984 U.S. cities with over 10,000 inhabitants in 1930. The following groups were mapped:

- manufacturing cities (2 levels)
- retail centers
- diversified cities
- wholesale centers
- transportation centers
- mining towns
- university towns
- resort and retirement towns.

Hickey, Richard J.; David E. Boyce; Evelyn B. Harner; and Richard C. Clelland. *Ecological Statistical Studies on Environmental Pollution and Chronic Disease in Metropolitan Areas of the United States.* RSRI Discussion Paper Series No. 38 (Philadelphia: Regional Science Research Institute, 1970).

No city classification or factor analysis is performed here. This study performs regression analyses to test the hypothesis that environmental pollution is a significant factor in public health. The results of the analysis are not inconsistent with the hypothesis. The study provides a useful inventory of pollutants which affect public health and related data sources.

Hoch, Irving. "Income and City Size." *Urban Studies* (Vol. 9, 1972): pp. 299-328.

This study deals with the causes and effects of city size. Five hypotheses are presented:

"(1) Large cities have become large because they have had 'natural' competitive advantages; this has been manifested in higher wages than offered at alternative locales, with a consequent inflow of labour, and hence, of population.

(2) Of course, item (1) refers to a disequilibrium situation, and at equilibrium, real wages everywhere will be the same.

(3) Money income per capita tends to increase city size; part of the increase can be explained by increasing cost of living, in terms of a conventional cost of living index. I believe the residual can be explained by greater net negative non-pecuniary returns with city size.

(4) The tendency toward equality of real incomes everywhere does not imply this is the best of all possible worlds, for improvement of institutional arrangements may better internalize externalities than compensatory payments for negative non-pecuniary returns.

(5) It is likely that such improvement will increase city size."

These hypotheses are supported with theoretical discussion and empirical evidence is presented in the form of statistics relating to (1) money income and deflated money income with city size; (2) rent, transportation, and congestion costs; (3) environmental quality; (4) health and welfare effects; and (5) positive effects of city size. A discussion of policy implications of relationships to city size concludes the paper.

Hughes, James W. *Urban Indicators, Metropolitan Evaluation, and Public Policy* (New Brunswick, New Jersey: Rutgers University, Center for Urban Policy Research, 1973).

The central purpose of this research is to analyze the evolution of "human spatial organization," thus facilitating more effective long range planning and policy formulation. "A major objective...is to determine whether metropolitan areas are converging toward similar patterns of residential differentiation. It is hypothesized that large metropolitan formations, drawn from wide ranging settings, converge toward a unique final steady state independent of sociocultural origin.

In order to test this hypothesis, a small group of diverse metropolitan observations were selected for analysis. Selection of sample metropolitan areas was by factor analysis of 24 demographic measures for the U.S. and Canada. The primary components yielded by the analysis were socioeconomic status, stage in life cycle, race, segregation and ethnicity.
Hughes notes the first three of these are almost identical to the Shevky factors (Shevky: 1955). This is neither proof of the method nor coincidence, since both studies utilized demographic data from the Census.

Jonassen, Christen T. and Sherwood H. Peres. Interrelationships of Dimensions of Community Systems: A Factor Analysis of Eighty-two Variables (Columbus: Ohio State University Press, 1960).

The major obstacle to application of factor analysis to cities is the enormous resources required to collect data on a large number of variables for a large number of cities. This study looks at 82 variables for 88 cities (1948-1953).
Observations comprised the 88 counties of Ohio--counties were used in order to include the rural factor. The following factors resulted.

Factor A: Urbanism: bi-polar factor with urbanism at one end and agrarianism at the other. The relevant variables represent wealth, power, education, professional abilities, vitality, productiveness and commercialism. Also high crime, delinquency and lack of social control. 51 of 82 variables load ≥ .30.

Factor B: Welfare: 16 variables load ≥ .30. Some of the more interesting:

child neglect -.77
educational sacrifice .62
crime -.27
mental illness -.57

technical illiteracy -.40
educational effort .53
social control .25

Factor C: Influx: high loadings on migration population change, dwelling newness, population change, retail sales gain and population mobility. Negative on commercial activity, wholesale trade, poverty. Influx is in middle-class suburbs and outlying areas.

Factor D: Poverty: loads highest on poverty variable.

Factor E: Magni-Complexity: "associated with large size and complex social and economic organization." High positive loads on economic base, population density, population size, relief expenditure, educational plant size, governmental complexity, hetrogeneity, wholesale trade, migration population change and retail sales gain.

Factor F: Educational Effort: high loadings on educational expenditure and educational status variable but not on wealth.

Factor G: Proletarianism: Highly negative on wealth differential, highly positive on unskilled workers.

Jones, Kenneth J. "Problems of Grouping Individuals and the Method of Modality." Behavioral Science (Vol. 13, Nov. 1968): pp. 496-511.

This article describes methods used in psychology, anthropology, and other sciences to select logical subsets from groups of observations, including three approaches: frequency curve resolution, factor analysis (taking factors as groups, which results in splits), and hierarchical clustering (using distance functions). He points out many weaknesses of these methods. For example, the use of frequency curves is practical only for one dimensional problems; factor analysis does not yield groups, but gives loadings to individuals proportional to distances from the origin on the factors; and distances are based on arbitrary interval scales. Method of rotation to achieve maximum modality is presented. This method is used in Jones & Jones, "Toward a Typology of Cities."

Jones, Kenneth J. and Wyatt C. Jones. "Toward a Typology of American Cities." Journal of Regional Science (Vol. 10, No. 2, August, 1970): pp. 217-224.

This study attempts to develop a classification of cities from factor descriptions provided by Hadden and Borgatta. Scope was reduced from 674 cities (over 25,000) to 315 cities (≥50,000). Second order principal components analysis was performed on Hadden and Borgatta's 12 factors to establish orthogonality and adjust factors to the new sample. Four factors resulted. Factor scores were rotated using a modality criterion to identify clusters. The modality criterion was to maximize the bimodality of the distribution as measured by kurtosis. Only two factors indicated statistically significant bimodality: (1) urban growth and development, and (2) socioeconomic status. An arbitrary cutting point of one standard deviation from the mean classified 52 cities among the four types described by these two bi-polar factors (one-sixth of cities). Graphical and descriptive interpretation of these groups indicates the typology is accurately descriptive, if not efficient.

Kass, Roy. "A Functional Classification of Metropolitan Communities." Demography (10:3, August 1973): pp. 427-445.

The author contends that although previous functional classifications have recognized multiple specializations of cities, each has required the specialization functions to be considered one at a time. This paper presents a means for cluster analysis which considers multiple dimensions simultaneously.

The SMSAs of 1960 were initially described in terms of location coefficients for each of eight industrial groups. "A community's location coefficient for an industry is the ratio of the proportion of the community's labor force employed in that industry to the corresponding proportion in the national labor force." The eight industrial groups utilized were (1) durable goods manufacturing, (2) nondurable goods manufacturing, (3) retail trade, (4) wholesale trade, (5) transportation, (6) finance, (7) education, and (8) public administration.

Given these eight functions, a correlation matrix of the communities' function profiles was developed. Pairs of communities and clusters were then identified based on relative correlations between communities.

Keeler, Emmett and William Rogers. A Classification of Large American Urban Areas. (National Science Foundation, May 1973).

The 125 largest American metropolitan areas are classified by this research, with emphasis on growth and decline of areas, social and racial inequality, government and political structure, and city-suburban relationships. Fifty-three related measures are transformed by factor analysis into eight factors: manufacturing, income inequality, ghetto-suburb contrasts, growth, age of population, unemployment, density and education.

"Two types of results are presented: a complete list of the cities divided into ten clusters, where each cluster can be characterized by its most representative city--Columbia, San Diego, South Bend, Knoxville, Dallas, Worcester, Cleveland, Oxnard, San Antonio, and Pittsburgh; and a tree of the cluster types where the cities are divided into 4, 6, 8, 10 and 14 clusters... Although there are no regional variables per se, the clusters generally represent geographic regions, since regions tend to share a common history and economic development, which is reflected in the data."

King, Leslie J. "Cross-Sectional Analysis of Canadian Urban Dimensions: 1951 and 1961." *The Canadian Geographer* (10:4, 1966): pp. 205-224.

This study is performed using factor analysis and is appropriately preceded by a discussion of the questionable value of using estimated communalities in factor analysis and several aspects of factor analysis concerning stability of a factor structure over time.

Analysis encompasses 106 cities which in 1951 had populations in excess of 10,000. More than 50 economic, demographic, social and locational characteristics were included.

The methodology included principal component analysis and a grouping algorithm discussed in J.H. Ward (1963).

Primary Factors:

1951: youthfulness of the female population, socio-economic character of cities outside of Quebec, frontier location and economic orientation, small city manufacturing, suburban occupational and housing structure, and metals employment and related housing conditions.

1961: urban textile manufacturing, Quebec population structure, metropolitan socioeconomic structure, high socioeconomic residential status, service role of older and comparatively isolated cities, and urban depression.

Grouping showed regional contiguity in both years.

King, Leslie J and Douglas Jeffrey. "City Classification by Oblique-Factor Analysis of Time Series Data." *City Classifiction Handbook*

Time series-cross section data are analyzed to demonstrate the capabilities of factor analysis in discerning causal factors. Urban economic activity is monitored through bimonthly levels of unemployment over a period of over four years for a set of 71 metropolitan areas in the eastern United States. Cities were classified by seven oblique factors--each group comprising those cities which scored highest on a given factor. Groupings showed a regional orientation and supported both the export-base theory and the industrial mix theory in different cases. As the initial factors were oblique, a second-order factor analysis looking at correlations between the factors was possible. Two second-order factors were developed. These were not highly correlated and were thought to reflect the "basic differences in the type of short-term economic forces affecting megalopolis, as opposed to non-megalopolis.

Kneedler, Grace M. "Economic Classification of Cities." *The Municipal Yearbook 1945*, ed by Clarence E. Ridley and Orin F. Nolting (Chicago: The International City Managers Assoc., 1945): pp. 30-38.

Cities of over 10,000 in the U.S. are classified using data from the 1940 Census of Population and the 1939 Census of Manufacturers.

Classification of cities by metropolitan status included five categories:

I independent-city is not a part of a United States Census metropolitan district.
C central - central city in a metropolitan district.
D a dormitory or residential suburb.
B a balanced suburb
E an employing suburb

The three types of suburbs were distinguished by an employment-resident ratio.

Classification of cities by rent level in metropolitan districts included characterization of suburbs by four rent levels relative to the average for the metropolitan area: low, intermediate, high, or exclusive.

Classification of cities by major economic base was by employment ratios among 11 categories.

These three classifications were developed independently.

Lampard, Eric E. "The Evolving System of Cities in the United States: Urbanization and Economic Development." *Issues in Urban Economics*, ed. by Harvey S. Perloff and Lowdon Wingo, Jr. (Baltimore: Johns Hopkins Press, 1968): pp. 81-139.

This paper is an investigation of the role of distance and spatial location in the urbanization of the U.S. An elaborate discussion of land use and location theory is followed by an analysis of relevant aggregate statistics. Conclusion: "The elements of structure and stability to be found in the evolving system of cities are broadly consistent with the notions of central place theory." (Central place theory is a theory of urban markets and services).

Marshall, John V. "City Size, Economic Diversity, and Functional Type: The Canadian Case." Economic Geography (51:1, Jan. 1975), pp. 37-49.

The primary purpose of this study was to test alternative measures of economic diversity and to relate economic diversity to city size and functional type. The study covers all incorporated municipalities in Canada with populations greater than 10,000 in 1961. Two measures of economic diversity were tested: the mean absolute index and the Gini index, of which the Gini index performed better. Interrelationships between size, diversity and functional type were examined by grouping cities (108) according to correlations based on 20 industry labor force measures. The study concludes that "very definite relationships between functional type and degree of diversification" are present.

Meyer, David R. "Classification of U.S. Metropolitan Areas by Characteristics of their Non-White Populations." City Classification Handbook

A classification of cities from a non-white perspective is developed by factor analysis using 40 variables for each of the 145 SMSAs in 1960 for which the U.S. Bureau of the Census compiled separate data for non-whites. Seven independent dimensions were determined: socioeconimic status of black Americans, status differentiation of non-white groups, stage in life cycle, old housing and high unemployment in northern industrial cities, multifamily housing in the northeast, the black urban hierarchy, and service workers in farming states. The results proved difficult to interpret. However, a sequential classificatory process was utilized to group the cities as follows:

Group	Subgroup
Northeastern	Northeastern
Southern	Texan Deep South North Carolina-Florida
Small Northern Industrial	Declining Industrial Young Family Industrial

Moser, C.A. and Wolf Scott. British Towns: A Statistical Study of Their Social and Economic Differences (London: Oliver and Boyd, 1961).

The objective of this research was to describe and classify British towns on the basis of social, economic and demographic characteristics. For the 157 towns in England and Wales with populations of 50,000 or more in 1951, sixty variables measured the following characteristics: population size and structure, population change, households and housing, economic character, social class, voting, health, and education. Principal component analysis was utilized to extract underlying factors, resulting in the following four components:

I. Social Class
II. Population Change Between 1931 and 1951
III. Population Change Between 1951 and 1958
IV. Overcrowding

The first component indicates the occupational structure; this factor explained the largest proportion of variation between towns: 30.1%. The second and third factors are measures of a town's stage of development and are highly correlated with certain characteristics of population and land use. The fourth component is related to housing density and other measures of housing quality.

Classification of the cities was achieved by plotting the cities against two factors at a time, making arbitrary assignments in marginal cases. Three categories and fourteen groups within those categories were identified.

Muth, Richard F. "Differential Growth Among Large U.S. Cities," Papers in Quantitative Economics, ed. by James P. Quirk and Arvid M. Zarley (Lawrence: The University of Kansas, 1968), pp. 311-355.

This paper is not directly relevant to city classification. The basis of the research is an exploration of growth in large cities. Hypotheses are tested with respect to migration patterns and employment elasticities.

Nelson, Howard J. "A Service Classification of American Cities." Economic Geography (31:3, July 1955): pp. 189-210.

"From material in the 1950 Census of Population, 897 areas of over 10,000 population were classified according to concentrations of employment into nine service groups."

manufacturing
retail trade
professional service
transportation & communications
personal service
public administration
wholesale trade
finance, insurance and real estate
mining.

"Concentrations were recognized if they were one or more standard deviations from an established national mean of employment in the various fields of service. Cities not high in any service were grouped together under a "diversified" class. A city could have recognized concentrations in more than one service."*

Nelson, Howard J. "Some Characteristics of the Population of Cities in Similar Service Classifications." Economic Geography (33:2, April 1957): pp. 95-108.

This article expands on the previous analysis described in "A Service Classification of American Cities," (Nelson, 1955). The first part of the analysis explores possible correlations between the functional characteristics of cities and socioeconomic characteristics such as population change, age, educational attainment, labor force participation, etc. An array of averages of selected measures of social attributes for each of the functional classifications determined in the previous study provided a framework for the analysis. Conclusion: "It is evident that variations in economic and social qualities of American cities are, in part at least, correlated with the function or service class to which a city belongs."

An investigation of the influence of location on social and economic attributes of cities in the same functional classification was based on a disaggregated version of a table which was subdivided to represent west, north central, northeast and south. Conclusion: "Generally, the characteristics of a region tend to be impressed on all classes of cities within an area, without, however, obliterating the large variations attributable to the service class."

---

*From Nelson, 1957.

Papageorgiou, G.S. "A Comparative Analysis of Fifteen Metropolises." Ekistics (32:188, July 1971): pp. 4-11.

Fifteen metropolises of the world have been characterised and compared in terms of area, population and density. They are also ranked with respect to groups of variables (apparently chosen based on qualitative analysis) representing (1) housing conditions, (2) central area size, density, (3) traffic (cars in relation to arterial roads, area, population), (4) public transport, and (5) economy. An overall ranking is calculated for each city based on sum of ranks in each of the five above categories. (summed up)

The 15 cities:

Athens
Vienna
Stockholm
New York
Detroit

Toronto
Bangkok
Tokyo
Madrid
Paris

London
Rio de Janeiro
Delhi
Karachi
Melbourne

Pidot, George B., and John W. Sommer. Modal Cities. EPA-600/5-74-027, October 1974.

"Modal cities are representative cities based on a specific set of criteria. Using principal components analysis, 224 U.S. SMSA's were examined in terms of 48 selected variables. This analysis yielded 14 dimensions, of which 7 explained 67% of the variance.

The 224 cities were then grouped using a method that minimizes the differences among cities within a group and maximizes the differences across groups. This procedure allowed for a confident selection of 9 modalities of the U.S. metropolitan system. Each city fell into a modality and was ranked relative to its distance from the mean. The two cities closest to the mean were taken as representative of that group. One unforeseen result of this research was the distinct regional character of the different groupings."

Price, Daniel O. "Factor Analysis in the Study of Metropolitan Centers." Social Forces (20:4, May 1942): pp. 449-455.

"The problem which give rise to this explanatory study was an analysis of the occupational characteristics and their social and economic correlates of the 93 cities with populations of 100,000 or more in 1930." The Thurstone method of factor analysis is utilized to find the fundamental dimensions of the system described by fifteen variables:

population
percent of gainfully employed in nonservice industry
sex ratios (males per 100 females)
percent increase in population, 1920-1930
median monthly rental of nonfarm homes
percent of total population unemployed
number of censal years in 100,000 class
percent of total population between ages 15 and 50
percent of population 10 years old and over
  gainfully occupied
percent of gainful workers which are female
median size of family
wholesale trade per capita
retail trade per capita
wages per wage earner
percent of employed making income tax returns.

Both oblique and orthogonal rotations were performed, however, the set of oblique factors proved uninterpretable. The final solution was a set of four orthogonal factors: "degree of maturity, the extent to which a city is a service center, the level of living within a city, and the per capita trade volume of a city." The cities were ranked according to each factor separately.

Ray, D. Michael. "The Spatial Structure of Economic and Cultural Differences: A Factorial Ecology of Canada." The Regional Science Association Papers (Vol. 23, 1969): pp. 7-23.

"The variation of economic and cultural characteristics across a national territory tends to exhibit three spatial patterns: a national heartland-hinterland pattern, an urban hierarchy pattern, and an intermetropolitan axis pattern. This paper examines the nature of these three spatial patterns in Canada, using multivariate procedures that have been applied in factorial ecologies of cities. The data anlayzed are 76 socio-economic variables and spatial locators for each of the 229 census counties in Canada, 1961. The variables selected stress the three phenomena of general interest in Canada: the extent of interregional disparities in economic development, the importance of cultural differences, and the contribution of foreign direct-investment to manufacturing activity. The variables are factor analysed to identify the underlying socio-economic dimensions and the factor scores for the counties on each dimension are grouped by distance and direction from the heartland. The significance of the heartland-hinterland and case patterns is tested by analysis of variance of the grouped factor scores. The results of these analyses offer a succinct summary of the spatial structure of economic and cultural differences and provide descriptive evidence of the prominence of the three spatial patterns in Canada."

Sacks, Seymour and Robert Firestine. "Dimensions and Classification of British Towns on the Basis of New Data." City Classification Handbook

This analysis is based on the Moser and Scott study, but includes new data from 1966 and certain methodological changes. The relationships of land uses to an urban hierarchy and to each other are investigated by correlating property value densities between land uses and between city types. City types included central cities (48), non-central cities (124), London metropolitan area (49), non-London area (123), central cities ILEA (47), non-central cities less London area (76). Land use types include domestic, shops, offices, other commercial, industrial, crown, and other. Comparison of land use among city types (which were not determined by factor analysis) not only demonstrated the existance of urbanization but the character of urbanization in term of value densities.

Shevky, Eshref and Wendell Bell. Social Area Analysis: Theory, Illustrative Application, and Computational Procedures. Stanford Sociological Series No. 1 (Stanford University Press, 1955).

A system of urban typology is described and demonstrated. "In order to delimit types of urban census tract populations, it makes use of dimensions defined as social rank, urbanization, and segregation. The social rank index is constructed from census tract measures of occupation and education; the urbanization index from measures of infertility, women in the labor force, and multiple family dwelling units; and the segregation index from measures of spatially concentrated ethnic groups. Arbitrary cutting points were established on the index values, and census tracts with similar score patterns were grouped together as types and designated as social areas."*

It appears the indices were constructed by arithmetic means of the relevant indicators--no attention being given to the optimal weight of the measures from which the indices were derived.

Sutton, Richard J.; John Korey; Steve Bryant; and Richard Dodson. "American City Types: Toward a More Systematic Urban Study," Urban Affairs Quarterly (9:3, March 1974): pp. 369-401.

A critical assessment of the value of previous efforts to classify cities is followed by a description of a method proposed for developing more accurate and useful classifications. "The method suggested here for constructing city types uses a combination of techniques--principal components analysis, the derivation of component scores, hierarchical clustering, a refinement clustering procedure, multiple discriminant analysis, and analysis of the group centroids to identify city group differences. Additionally, the cities are described in a theoretically relevant way.
Nine factors were identified, of which seven were important in group discrimination. The eight chosen groups are described by the final seven factors: socioeconomic status; age and ethnicity; nonmanufacturing labor force; rurality; residential expansion; manufacturing expansion; and private regardingness.

Thorndike, E.L. Your City. (New York: Harcourt, Brace & Company, 1939).

Thorndike uses a data base of thirty-seven measures to describe 310 cities with populations greater than 30,000 with respect to "general goodness of life for good people." Three indices are also constructed from these measures to reflect 1) general goodness of life for good people, 2) desirable personal qualities of a population, and 3) per capita private income. Each of the thirty-seven basic measures is expressed in deviations from the mean, and these three indices are constructed by weighting the individual measures (the derivation of these weights is not revealed). Thorndike has examined the intercorrelation between the three indices and between indices and basic measures to ascertain causal relationships between quality of life and the characteristics of cities. The study concludes that "at least four fifths of the differences of cities in goodness is caused by the personal qualities of the citizens and the amount of their incomes."

---

*From Van Arsdol, "An Investigation of the Utility of Urban Typology," Pacific Sociological Review (Vol. 4, Spring 1961): pp. 26-32.

Thorndike, E.L. *144 Smaller Cities* (New York: Harcourt, Brace & Company, 1940).

This is a sequel to *Your City*. Data for cities between 20,000 and 30,000 population in 1930 has been gathered using the same measures as the previous volume. The analytic process of *Your City* is repeated and results in very similar conclusions, which are presented as suggestions for the small city. Thorndike concludes "good people make a city good" and suggests, "the surest way to have good people is to breed them."

Van Arsdol, Maurice D. Jr.; Santo F. Camilleri; and Calvin F. Schmid. "The Generality of Urban Social Area Indexes." *American Sociological Review* (23:3, June 1958): pp. 277-284.

"This research is concerned with the empirical generality of the Shevky system of social area analysis. The factor analysis tests of hypotheses show that the present form of the Shevky system has high generality for the cities included in this study.* Although the Shevky measures do appear to differentiate social rank, urbanization, and segregation factors in the spatial patterning of cities, minor variation is observed in the internal structure of these factors. It is not known whether the variation from the Shevky system has occurred by chance. Alternative models of the measures and dimensions were described for some of the cities, but due to the ex post facto nature of this portion of the analysis, no priority can be claimed for the alternative models over the Shevky model."

---

*Akron, Atlanta, Birmingham, Kansas City, Louisville, Minneapolis, Portland, Providence, Rochester and Seattle. Data from 1950 Census.

Ward, Joe H. Jr. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* (58:301, March 1963): pp. 236-244.

"A procedure for forming hierarchical groups of mutually exclusive subsets, each of which has members that are maximally similar with respect to specified characteristics is suggested for use in large-scale ($n > 100$) studies when a precise optimal solution for a specified number of groups is not practical. Given n sets, this procedure permits their reduction to n-1 mutually exclusive sets by considering the union of all possible n(n-1)/2 pairs and selecting a union having a maximal value for the functional relation, or objective function, that reflects the criterion chosen by the investigator (minimizing increase in sum of squared errors is proposed). By repeating this process until only one group remains, the complete hierarchical structure and a quantitative estimate of the loss associated with each stage in the grouping can be obtained. A general flowchart helpful in computer programming and a numerical example are included."

Wilson, John Oliver. "Quality of Life in the United States - An Excursion into the New Frontier of Socio-Economic Indicators." *The Quality of Life Concept* (Washington, D.C.: Environmental Protection Agency, 1973).

This article describes one more approach to quantifying quality of life. The analysis is based on data for 72 measures and 50 states. Both classical factor analysis and principal components analysis were utilized in development of eight separate indicators. These indicators, suggested by a set of goals for the nation formulated by the Commission on National Goals, are as follows: individual status, individual equality, education, economic growth, technological change, agriculture, living conditions, health and welfare. Classical factor analysis and principal components analysis yielded similar results. States were ranked according to each of the eight factors separately.

APPENDIX C

BMD COMPUTER PACKAGE DESCRIPTION

The 1975 version of the Biomedical Computer Programs (BMDP) package was utilized in the extensive statistical analysis performed under this contract. The BMD package includes a large number of programs which deal with data description, frequency tables, multivariate analysis (factor analysis, clustering, canonical correlation, etcetera), regression, analysis of variance, and other special techniques. An excellent description of the programs and their uses is available in manual form* from:

University of California Press
2223 Fulton Street
Berkeley, California 94720

For the work under this contract, the BMD package was purchased on magnetic tape and installed at Optimum Systems, Inc. (OSI).**

Programs Used

BMD programs used in the present research include:

P1D: Simple Data Description
P3D: t-test and T^2 routine
P2D: Frequency Count Routine
P2M: Cluster Analysis on Cases
P3M: Block Clusterings
P4M: Factor Analysis
P1R: Multiple Linear Regression.

The application of each to this research will be described in turn.

Program P2D: Frequency Count Routine

As part of the data cleaning process, a histogram was generated for each variable to isolate obvious outliers. This program provides a comprehensive set of descriptors for each distribution to aid the data screening process.

---

*Price: $10.

**An earlier version of BMD was already available at OSI, but proved to be inadequate for this research. The tape is available from BMD Program Librarian, Health Sciences Computing Facility, AV-111 Health Sciences Center, University of California, Los Angeles, California 90024. Cost is $76 or $86 depending on tape format options selected (see manual). Programs are written in IBM Fortran G and H. Instructions are provided for implementation on IBM hardware with core of at least 128K. The technical support at OSI (800-638-8743) and the BMD program librarian can provide assistance in setting up the package.

Observations shown to be outliers by these statistics and determined to have unreasonable values for the given variable were checked and either corrected or removed from the data base.

Program P3D: t-Test and T^2 Routine

This program was used in testing the significance of differences between SMSA groups according to environmental measures. t-Statistics were generated to compare group means for each of seven variables, and F-statistics were calculated to measure the overall difference between groups based on the seven environmental measures. The result was a t-statistic measuring the discriminating ability of each pair of groups on each environmental measure and an F-statistic measuring the discriminating ability of each pair of groups on the environmental variables as a whole.

Program P2M: Cluster Analysis on Cases

Program P2M was used to group similar SMSAs. Factor scores describing each SMSA in a geometric space defined by the factors were input from a save file created by P1D. A "nearest neighbor" clustering procedure is used to form groups, progressing from a large number of single observations to a single group of all observations. The program provides no criterion for stopping at some point in the grouping process--the output includes a description of the group formed at each step and a tree diagram depicting all subgroups. See the annotated output of Appendix II-D.

A second cluster analysis program was used to supplement BMD program P2M. This auxiliary program, written by Stephen Tweedie and modified by Kenneth Mayer of Syracuse University, and further modified by Richard Tabors and James Hudson of USR&E, provided a number of summary statistics unavailable through the BMD program. Specifically, it enabled analysis of the remaining levels of within-group and between-group variance at each step in the clustering process. In several tests, the two programs produced identical clusters on the full set of 262 cities, and the BMD program was used for basic runs because of better output formats. The auxiliary program is written in Fortran IV, G Level (IBM release 18), and is available from the project, in either its well-documented unmodified form or its modified form used in the research.

Program P4M: Factor Analysis

This program performs a factor analysis to generate factors defined in terms of the input variables and factor scores describing individual observations in terms of the new dimensions (factors). A wide variety of options are available with respect to the type of input, method of extracting initial factors, and method of rotating the factors to an optimal solution. Raw data was input to the program in our applications (following cleaning and tabulation). Initial factors were determined by principal components analysis and rotation of the factors was by the varimax technique in general. Experimentation in varying these techniques seemed to indicate the results were relatively robust with respect to alternative methods although minor changes were observed.

Four steps in the factor analysis program are important in this research: (1) initial factor extraction; (2) factor rotation; (3) factor score generation; and (4) creation of a save file. These processes are delineated in the annotated output of Appendix II-D. Factor scores generated by this program are input directly to the clustering program.

Program P1R: Multiple Linear Regression

This program is a standard regression routine using ordinary least squares. It was used to analyze the relationship between the stage 1 factors and individual environmental quality indicators, as described in the main body of the text.

Using the Package on OSI

The OSI computer runs IBM OS with the Wylbur Editor. Essentially, this means that the time sharing system allows editing and preparation to be performed at the console, with programs submitted to run in a batch environment. BMD programs are individually accessed through IBM Job Control Language, and submitted as batch jobs. When the processing is complete, the output can be printed at OSI's Washington Data Center, at regional centers (including EPA computer facilities), or at the console. Because the output from BMD is formatted for batch and contains significant extraneous information, printing at the console can be slow and unwieldy. However, successful job completion can be assured before off-line printing by using the Wylbur editor system to search job stream output.

Comments on the Use of BMD and OSI for this Research

The BMD package on the OSI system provided an effective tool for this research. The package allows free format input and a wide variety of options with respect to input, methodology, and output. Individual programs are extremely powerful and are well-documented. The greatest weakness of the package was found in combining output from a group of programs to form the input for further analysis. This was possible by combining save files through the Wylbur editor system. The combination of the systems is efficient if not interactive.

APPENDIX D

ANNOTATED COMPUTER OUTPUT: FACTOR ANALYSIS AND CLUSTERINGS

This appendix is included to provide examples of the factor analysis and clustering procedures used in this research. The output from two consecutive programs is included in annotated form for the analysis performed on socio-demographic data. The first program is BMDP4M, Factor Analysis, which estimates factor scores for each SMSA on the important factors. The second program, BMDP2M, applies a clustering algorithm to these factor scores to derive groups of SMSAs which are similar according to the derived factors.

```
BMDP4M - FACTOR ANALYSIS - DOUBLE PRECISION VERSION          PROGRAM REVISED JULY 7, 1975
HEALTH SCIENCES COMPUTING FACILITY                           MANUAL DATE -- 1975
UNIVERSITY OF CALIFORNIA, LOS ANGELES


 PROGRAM CONTROL INFORMATION

          /PROBLEM     TITLE IS 'SCH-EI-TESTMAN'.
          /INPUT       VARIABLES ARE 40.
         FORMAT = '[illegible]'.
              CASES ARE 262.    TYPE IS DATA.
          /VARIABLE
       NAMES = [illegible]
        [illegible]
        [illegible]
          [illegible]
        [illegible]
          [illegible]
        LABEL=1,2.
          MISSING [illegible]
        /TRANSFORM [illegible]
          [illegible]
            [illegible]
        /SAVE UNIT IS 4. NEW. CODE IS SCHDATA. CONTENT IS DATA.
          /END


 PROBLEM TITLE . . . . . . . SCH-EI-TESTMAN

 NUMBER OF VARIABLES TO READ IN. . . . . . . . . .        40
 NUMBER OF VARIABLES ADDED BY TRANSFORMATIONS. .           8
 TOTAL NUMBER OF VARIABLES . . . . . . . . . . . .        48
 NUMBER OF CASES TO READ IN. . . . . . . . . . . .       262
 CASE LABELING VARIABLES . . . . . . . . . . . . .      X(1)          X(2)
 LIMITS AND MISSING VALUE CHECKED BEFORE TRANSFORMATIONS
 BLANKS ARE. . . . . . . . . . . . . . . . . . . .     ZEROS
 INPUT UNIT NUMBER . . . . . . . . . . . . . . . .         5
 REWIND INPUT UNIT PRIOR TO READING. . DATA. . .          NO

 INPUT FORMAT
        [illegible]


 INTERPRETIVE TRANSFORMATIONS ARE
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .
        [illegible]     = [illegible]    *   [illegible]       .
        [illegible]     = [illegible]    /   [illegible]       .


 VARIABLES TO BE USED
             3  AGE         6  [illegible]     7  FA          8  [illegible]      9  [illegible]
            10  [illegible] 11  [illegible]   12  [illegible] 18  PCC            19  PCS
            21  [illegible] 22  [illegible]   23  [illegible] 24  [illegible]    25  [illegible]
            2[illegible] EPT 28  [illegible]  29  DR          30  AUTO           31  TD
            3[illegible] [illegible] 3[illegible] [illegible] 41 [illegible] 42 [illegible] 43 [illegible]
```

Input to BMD factor analysis program. Describes data, specifies method by which factor analysis will be performed and directs output to same file.

Initial output describes how the program interprets the input, beginning with data.

The program has the capability to perform simple arithmetic operations on the data and can [illegible] transformations on the raw data as input to the factor analysis. These are the new variables requested above.

Variables from the list of transformed variables and raw data to be input into the analysis.

```
            44  HSU            45  HSDU           46  EPH            48  HSGP

WEIGHT VARIABLE . . . . . . . . . . . . . . . . . . .
UNROTATED FACTORS ARE PRINCIPAL COMPONENTS.
NUMBER OF FACTORS IS LIMITED TO THE NUMBER OF EIGENVALUES GREATER THAN      1.000
VARIMAX ROTATION IS PERFORMED.
GAMMA . . . . . . . . . . . . . . . . . . . . . . . . .    1.0000
MAXIMUM NUMBER OF ITERATIONS FOR ROTATION . . . .     50
CONVERGENCE CRITERION FOR ROTATION. . . . . . . . 0.0000100
KAISER'S NORMALIZATION. . . . . . . . . . . . . . .     YES

REQUESTED OUTPUT AND SAVE FILE
   UNIT    =    9
   CODE    = SES[illegible]
   LABEL   =
   CONTENT =
           DATA          (INPUT DATA MATRIX AND FACTOR SCORES)
```

Choices from program options. The method in this case is described by the default options. Therefore, no related program input was necessary

Program interpretation of instructions for disposition of output.

```
DATA AFTER TRANSFORMATIONS FOR FIRST     5 CASES
CASES WITH ZERO WEIGHTS AND MISSING DATA NOT INCLUDED.

       CASE
LABEL       NUMBER        WEIGHT       3 AGE         4 PF          7 FR          8 NOR         9 PU          10 NMOB       11 ML
                      12 HTR        18 PCI        19 PCS        21 GINI       22 YW         23 MP         24 EMW        25 CPF
                      26 EPT        28 HC         29 DR         30 AUTO       31 TS         32 TP         36 DI         41 MBL
                      42 MCL        43 MOCC       44 MSU        45 MSDU       46 EPH        48 MISP
AMOL 45        1          1.00000      24.30000       4.40000    2444.00000       1.40000      66.30000      41.80000       2.40000
                          1.26000      -6.40000      -5.30000       0.37000    7516.00000    2513.00000      40.09000      28.67000
                         39.09000       7.90000       0.86000      91.30000      74.92000       4.33000      13.56000       0.05667
                          0.87737       6.59051       0.03532       0.01004       0.29000       0.25076
BHAN 80        2          1.00000      27.30000      14.20000    3056.00000       1.20000      63.60001      55.22000       3.21000
                          1.57962      -5.10000      12.20000       0.31000   11046.0000     3372.00000      56.91000      26.36000
                         34.20000       5.70000       0.65000      98.90000      76.70000      34.30000      25.40000       0.07900
                          0.77757       1.60000       0.06054       0.02245       0.31441       0.11233
ALBS 12M       3          1.00000      26.20000      23.00000    3361.00000       1.46000      65.50000      41.87000       3.42000
                          1.75000      24.80000      18.30000       0.39000    7975.00000    2463.00000      45.48000      27.91000
                         34.14000      13.70000       0.93000      91.30000      64.44000      14.40000      20.00000       0.13624
                          0.72057       9.46176       1.05477       0.00070       0.30162       0.24435
ALMA 167       4          1.70000      23.00000      23.10001    3116.00000       1.10000      75.30000      54.22000       3.34000
                          1.12000       7.40000       4.40000       0.32000   11944.0000     3356.00000      55.94000      29.34000
                         39.52000       4.20000       1.47000      82.00000      64.92000      21.21001      19.60000       0.03276
                          1.57000       0.50941       0.04761       0.01433       0.75241       0.10380
ALMA 208       5          1.00000      24.20000       9.40000    3324.00000       1.00000      94.20000      46.80000       3.24000
                          1.24000      21.20001      21.34000       1.37000    4627.00000    2063.00000      36.22000      29.14000
                         34.52000      10.30000       1.07000      69.70000      74.23000      11.12000      25.70000       0.02119
                          0.43251       0.63449       1.07713       0.03967       0.26423       0.28126

  NUMBER OF CASES READ. . . . . . . . . . . . . . . . . .     262
     CASES WITH DATA MISSING OR BEYOND LIMITS . .      31
        REMAINING NUMBER OF CASES . . . . . . . . . . .     231
```

A small sample of the data to be used in the factor analysis is presented
to allow a check on the input, transformations, variable names, etc.
Note 31 cases were eliminated from the data base due to missing data
for one or more variables (no limits were specified for any variable).

```
UNIVARIATE SUMMARY STATISTICS

                                                                            SMALLEST  FIRST                     LARGEST   FIRST
                                STANDARD      COEFFICIENT      SMALLEST     STANDARD CASE FOR      LARGEST     STANDARD CASE FOR
VARIABLE           MEAN         DEVIATION     OF VARIATION       VALUE        SCORE   SMALLEST       VALUE        SCORE   LARGEST

   3 AGE          27.47273        2.84571       0.104043       20.7000       -2.26    137          37.7000       3.4?     77
   6 DF           14.47610       11.09644       0.764425        0.9000       -1.23     83          48.7000       2.96    117
   7 FW         3150.51944      317.64150       0.100567     2557.0000       -1.84    141        4938.0000       5.66    137
   8 NBX           1.21342        0.49585       0.408635        0.3000       -1.84    195           3.8000       5.22    137
   9 PG           81.02987       11.37553       0.140734       48.2000       -2.87    194         102.0000       1.84    183
  10 [illegible]  50.61041        8.27223       0.161891       28.0700       -2.77    123          74.1200       2.34    205
  11 [illegible]   3.14671        0.19906       0.060081        2.6900       -2.47    235           4.2130       5.26    137
  12 FTM           1.04364        0.07577       0.071234        0.7100       -4.67    120           1.2420       2.33     36
  18 PCC          14.53676       20.66050       1.471717      -21.3000       -1.25    104         163.1000       5.34    162
  19 PCS          17.27443       16.13356       0.933977       -9.3000       -1.65     42         115.2000       6.07    110
  21 GINI          0.36431        0.03215       0.093441        0.2700       -2.30     33           0.4400       3.42    137
  22 [illegible] 9629.67200     1396.19821       0.144702     4761.0000       -3.49    137       14312.0000       3.36    163
  23 [illegible] 3055.61945      449.54421       0.147121     1482.0000       -3.50    137        4971.0000       4.26    153
  24 EMP          52.47623        5.75311       0.108548       25.1400       -4.86     73          60.6100       1.36    115
  25 [illegible]  26.29942        3.67527       0.138301       16.2000       -3.26    149          34.9000       2.37     18
  26 [illegible]  37.84108        3.62996       0.161896       22.1900       -4.36     73          44.7400       1.86    126
  28 HC            7.97403        3.55142       0.445371        3.7000       -1.25      7          33.0000       7.05    137
  29 [illegible]   1.09236        0.08376       0.793063        0.4489       -2.54    119           1.1700       3.32    224
  30 AUTO         65.54092        3.84205       0.058645       55.1000       -6.77    159          98.1200       2.49     24
  31 TG           78.44841        4.57530       0.093047       34.5200       -5.50    159          88.9400       2.36     18
  32 TP           16.54426        5.55356       0.333739        4.7300       -1.74     38          52.9400       6.54    159
  39 DI           21.65344        3.68038       0.170657       11.2000       -2.56    105          32.9000       3.06     59
  41 PHL           0.09733        0.09161       0.941235        0.0063       -1.06    165           0.3752       3.03    126
  42 [illegible]   0.73665        0.13747       0.186156        0.0363       -5.05     86           0.9213       1.34    244
  43 [illegible]   0.60043        0.09016       0.150854        0.0269       -6.37     86           0.7682       1.86     23
  44 [illegible]   0.09301        0.05304       0.570281        0.0044       -1.68     80           0.3746       5.31    163
  45 [illegible]   0.02575        0.02243       1.043255        0.0001       -4.91     73           0.2200       9.07    154
  46 EMM           0.31050        0.04196       0.144682        0.0324       -6.61     46           0.5530       3.76     63
  48 [illegible]   0.25329        0.06019       0.445130        0.0314       -1.75    159           0.3573       2.04     40

CASE NUMBERS ABOVE REFER TO DATA MATRIX BEFORE ANY CASES HAVE BEEN DELETED DUE TO MISSING DATA.
CASES WITH ZERO WEIGHTS ARE NOT INCLUDED.
```

Selected statistics describing the data are presented to facilitate the identification of observations which may be erroneous or unusual and to assist in the interpretation of factor analysis results.

CORRELATION MATRIX

		AGE 3	[illegible] 6	[illegible] 7	[illegible] 8	[illegible] 9	[illegible] 10	[illegible] 11	[illegible] 12	[illegible] 14	[illegible] 19	[illegible] 21	[illegible] 22	[illegible] 23
AGE	3	1.000												
[illegible]	6	0.277	1.000											
[illegible]	7	-0.502	0.133	1.000										
[illegible]	8	0.042	0.226	0.094	1.000									
[illegible]	9	-0.013	0.355	-0.044	0.023	1.000								
[illegible]	10	0.394	0.322	0.165	0.242	-0.206	1.000							
[illegible]	11	-0.580	0.094	0.744	0.304	-0.063	0.140	1.000						
[illegible]	12	0.494	0.137	0.021	0.263	-0.084	0.567	-0.044	1.000					
[illegible]	14	-0.242	-0.154	-0.044	-0.172	0.104	-0.480	-0.005	-0.225	1.000				
[illegible]	19	-0.292	-0.021	-0.174	-0.145	0.209	-0.582	-0.068	-0.304	0.563	1.000			
[illegible]	21	-0.132	-0.145	0.104	0.520	0.057	-0.315	0.128	0.050	0.204	0.063	1.000		
[illegible]	22	0.140	0.298	-0.304	-0.444	0.236	0.095	-0.222	-0.066	-0.077	0.245	-0.673	1.000	
[illegible]	23	0.321	0.296	-0.517	-0.371	0.333	-0.034	-0.433	0.034	-0.085	0.324	-0.480	0.922	1.000
[illegible]	24	0.341	0.044	-0.322	-0.204	-0.143	0.376	-0.125	0.502	-0.145	-0.136	-0.436	0.553	0.522
[illegible]	25	0.144	0.036	-0.405	-0.173	0.171	-0.362	-0.405	0.041	0.014	0.176	-0.344	0.404	0.447
[illegible]	26	0.394	0.067	-0.460	-0.236	0.072	0.216	-0.454	0.307	-0.076	0.024	-0.380	0.590	0.466
[illegible]	28	-0.413	0.122	0.043	0.544	0.132	-0.022	0.703	0.093	0.063	-0.073	0.506	-0.505	-0.008
[illegible]	29	0.323	-0.219	-0.281	0.247	-0.186	0.267	-0.212	0.229	-0.230	-0.303	0.176	-0.136	-0.136
AUTO	30	-0.393	-0.325	0.112	-0.440	-0.076	-0.409	0.107	-0.381	0.334	0.375	-0.208	0.265	0.247
[illegible]	31	-0.162	-0.342	0.427	-0.196	-0.071	-0.231	0.032	0.461	0.284	0.272	-0.036	-0.052	-0.095
[illegible]	32	0.275	0.326	-0.146	0.302	0.230	0.354	-0.054	0.262	-0.277	-0.221	-0.032	0.286	0.240
[illegible]	34	0.211	-0.198	-0.172	0.218	-0.021	0.058	-0.029	0.050	-0.188	-0.036	0.008	-0.136	-0.045
[illegible]	41	-0.124	-0.409	-0.207	0.340	0.061	-0.183	0.247	-0.006	0.055	0.040	0.445	-0.239	-0.253
[illegible]	42	-0.243	-0.554	0.207	0.136	-0.365	-0.148	0.203	-0.043	0.031	-0.065	0.173	-0.220	-0.360
[illegible]	43	0.122	-0.244	0.174	-0.135	-0.303	0.291	0.247	0.136	-0.162	-0.168	-0.332	0.113	-0.088
[illegible]	44	0.172	0.449	-0.320	-0.087	0.442	-0.131	-0.322	-0.044	0.030	0.250	-0.107	0.474	0.533
[illegible]	45	0.243	0.394	-0.282	0.104	0.344	-0.053	-0.211	0.094	0.027	0.107	0.108	0.248	0.365
[illegible]	46	0.133	-0.113	-0.253	0.024	0.049	0.078	-0.174	0.231	-0.073	0.029	-0.081	0.291	0.347
[illegible]	48	-0.419	-0.324	-0.114	-0.285	0.111	-0.485	-0.202	-0.603	0.502	0.591	0.344	-0.381	-0.088

		[illegible] 24	[illegible] 25	[illegible] 26	[illegible] 28	[illegible] 29	AUTO 30	[illegible] 31	[illegible] 32	[illegible] 34	[illegible] 41	[illegible] 42	[illegible] 43	[illegible] 44
[illegible]	24	1.000												
[illegible]	25	0.528	1.000											
[illegible]	26	0.465	0.825	1.000										
[illegible]	28	-0.437	-0.459	-0.528	1.000									
[illegible]	29	0.011	-0.084	-0.030	-0.111	1.000								
AUTO	30	-0.026	-0.093	-0.061	-0.126	-0.483	1.000							
[illegible]	31	0.344	-0.227	-0.067	0.012	-0.147	0.714	1.000						
[illegible]	32	0.181	0.256	0.247	-0.027	0.329	-0.793	-0.766	1.000					
[illegible]	34	-0.106	-0.125	-0.115	0.067	0.303	-0.236	-0.036	0.292	1.000				
[illegible]	41	-0.182	0.048	-0.114	0.196	0.450	-0.272	-0.129	0.277	0.422	1.000			
[illegible]	42	-0.252	-0.353	-0.317	0.198	-0.018	0.428	0.481	-0.582	0.062	0.066	1.000		
[illegible]	43	0.172	-0.288	0.023	-0.137	-0.054	0.348	0.306	-0.276	0.028	-0.244	0.763	1.000	
[illegible]	44	0.178	0.455	0.350	-0.202	-0.144	-0.331	-0.446	0.453	-0.031	-0.086	-0.585	-0.387	1.000
[illegible]	45	0.127	0.181	0.183	-0.026	-0.026	-0.336	-0.294	0.474	0.383	0.113	-0.360	-0.290	0.682
[illegible]	46	0.491	0.444	0.530	-0.147	0.025	-0.018	-0.055	0.162	0.066	0.114	-0.081	0.035	0.365
[illegible]	48	-0.447	0.078	-0.247	-0.011	-0.244	0.502	0.243	-0.496	-0.150	0.073	0.371	-0.229	0.056

		[illegible] 45	[illegible] 46	[illegible] 48
[illegible]	45	1.000		
[illegible]	46	0.208	1.000	
[illegible]	48	-0.037	-0.084	1.000

The correlation matrix is the basis for the factor analysis. Variables which show a high correlation here will appear in the same factor. Explanation of the factors is often facilitated by the examination of correlations since the relationships between variables are shown one by one here.

SQUARED MULTIPLE CORRELATIONS (SMC) OF EACH VARIABLE WITH ALL OTHER VARIABLES AND KAISER'S (1970) MEASURE OF SAMPLING ADEQUACY (MSA)

	SMC	MSA
TOTAL		0.74
3 AGE	0.84283	0.724
6 AF	0.81866	0.723
7 FA	0.83867	0.774
8 NOH	0.76953	0.736
9 PU	0.51875	0.715
10 NMOH	0.86667	0.833
11 PRH	0.86761	0.765
12 FTH	0.73381	0.842
18 PCI	0.42467	0.654
19 PCS	0.64765	0.788
21 HIHI	0.67767	0.632
22 TH	0.47832	0.574
23 HD	0.97642	0.466
24 EPH	0.97154	0.668
25 EPF	0.96426	0.453
26 EPT	0.99966	0.472
28 HC	0.69231	0.422
29 DH	0.62466	0.641
30 AUTS	0.44466	0.627
31 TO	0.42126	0.721
32 TH	0.44142	0.461
36 DI	0.40440	0.788
41 PHL	0.74661	0.645
42 HIU	0.67618	0.718
43 HDII	0.85089	0.623
44 HSU	0.73716	0.862
45 HSOU	0.56706	0.676
46 EPN	0.44226	0.827
47 HDSP	0.40633	0.744

These two measures indicate the overall correlation between each variable and the rest. An extremely low value in either column indicates the variable has little variance in common with any other subset of variables involved and will not combine with other variables to yield a factor.

[illegible]

[illegible]

For each variable, the communality is the proportion of the variance of that variable that can be explained by the seven dominating factors (defined arbitrarily here as those with eigenvalues greater than one). A low communality indicates the variable is not represented, or not represented strongly, in the selected factors.

[illegible]

In principal components analysis, the amount of variance explained by each factor is of great concern since a number of factors is chosen to represent the data set (according to the proportion of variance explained). In this case, the first seven factors each explain more than [illegible] of the total variance and together explain [illegible] percent of the total variance.

[illegible]

THE WEIGHTED [illegible] OF EACH FACTOR IS THE [illegible] FOR THAT FACTOR.

TOTAL WEIGHTED [illegible] IS [illegible] AS THE SUM OF THE [illegible] ELEMENTS OF THE [illegible] [illegible] MATRIX.

UNROTATED FACTOR LOADINGS (PATTERN)
FOR PRINCIPAL COMPONENTS

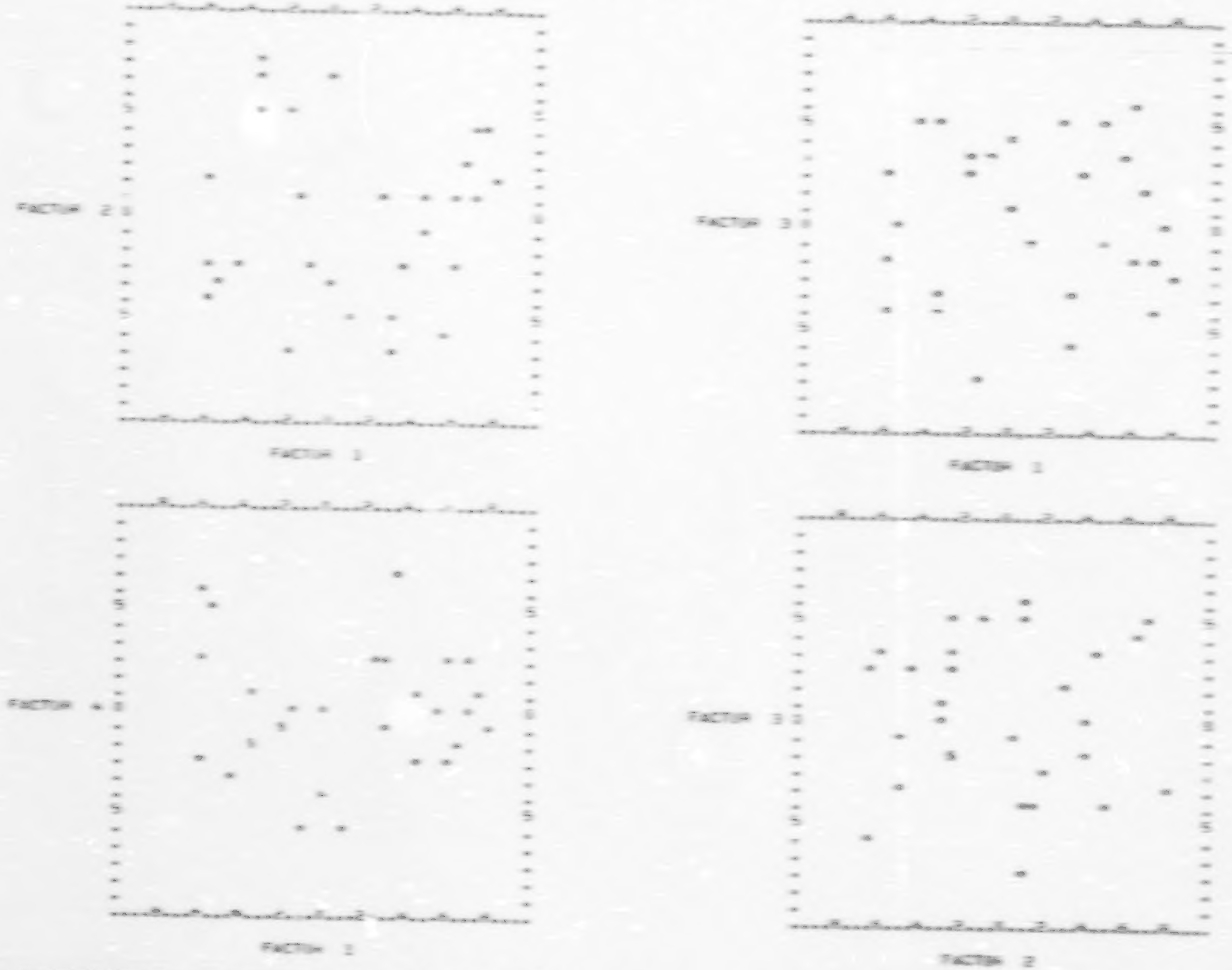
		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
AGE	3	0.586	-0.284	-0.193	-0.254	-0.176	-0.433	0.404
MF	6	0.366	-0.270	0.225	0.690	-0.075	-0.223	0.222
FS	7	-0.585	-0.242	-0.145	0.592	0.385	0.006	-0.344
HOH	8	-0.142	-0.083	0.256	-0.082	0.352	-0.196	0.362
PU	9	0.248	0.000	0.510	0.276	0.157	0.093	0.300
[illegible]	10	0.290	-0.641	-0.545	0.237	0.010	-0.019	0.037
[illegible]	11	-0.570	-0.295	-0.026	0.409	0.325	0.371	-0.032
FTW	12	0.324	-0.483	-0.339	-0.066	0.446	-0.333	0.255
MCC	14	-0.216	0.466	0.382	-0.073	0.286	-0.149	0.058
MCS	19	-0.015	0.625	0.466	0.011	0.195	0.051	0.239
GINI	21	-0.456	-0.365	0.522	-0.324	0.385	-0.306	0.081
TW	22	0.785	0.421	-0.152	0.294	0.436	0.372	0.127
WP	23	0.776	0.411	0.019	0.131	0.037	0.217	0.255
EPW	24	0.788	0.575	0.444	-0.041	0.393	-0.067	-0.115
EMF	25	0.670	0.276	0.135	-0.103	0.257	-0.076	-0.401
EPT	26	0.860	0.290	-0.233	-0.107	0.354	-0.081	-0.226
AC	28	-0.622	-0.447	0.296	0.257	0.416	-0.026	0.049
DA	29	0.110	-0.525	-0.000	-0.596	-0.263	0.123	0.022
AUTO	30	-0.367	0.791	-0.294	0.097	0.136	0.045	0.111
TD	31	-0.376	0.491	-0.379	-0.139	0.244	-0.297	0.301
TP	32	0.567	-0.593	0.313	0.023	-0.026	0.291	-0.110
DI	34	0.062	-0.384	0.073	-0.445	-0.054	0.305	0.423
PHL	41	-0.125	-0.298	0.366	-0.615	0.207	0.398	-0.036
HSU	42	-0.607	0.142	-0.443	-0.295	0.171	0.128	0.161
HOCC	43	-0.257	0.096	-0.747	0.009	0.183	0.235	0.245
HSU	44	0.623	0.062	0.546	0.225	-0.035	0.021	0.032
HSOU	45	0.464	-0.095	0.502	0.058	0.133	0.051	0.326
EPU	46	0.432	0.071	-0.072	-0.237	0.553	0.105	-0.258
HOGP	48	-0.376	0.609	0.476	-0.195	-0.040	-0.105	-0.106
	VP	6.723	4.484	3.836	2.772	1.767	1.392	1.326

THE VP FOR EACH FACTOR IS THE SUM OF THE SQUARES OF THE ELEMENTS OF THE COLUMN OF THE FACTOR LOADING MATRIX CORRESPONDING TO THAT FACTOR. THE VP IS THE VARIANCE EXPLAINED BY THE FACTOR.

This array expresses each of the 29 variables in terms of the seven major factors, where both variables and factors are standardized to mean zero and standard deviation one. For example,

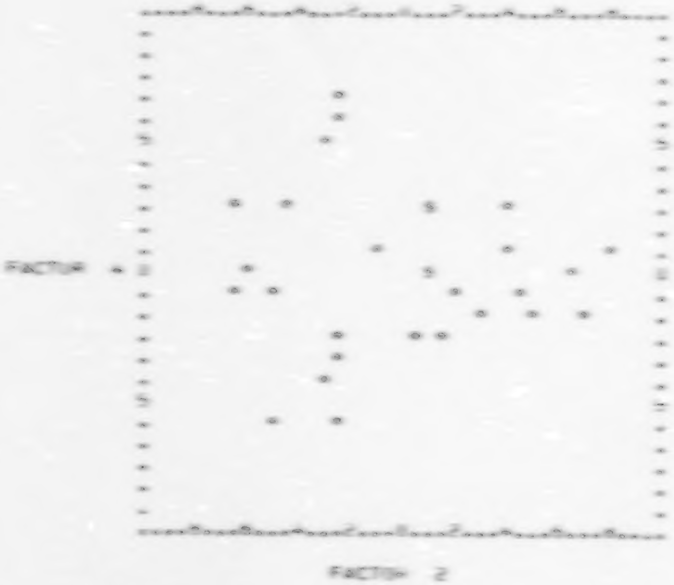
AGE = .586 x FACTOR 1 - .284 x FACTOR 2 - .193 x FACTOR 3 - .254 x FACTOR 4 - .176 x FACTOR 5 - .433 x FACTOR 6 + .404 x FACTOR 7

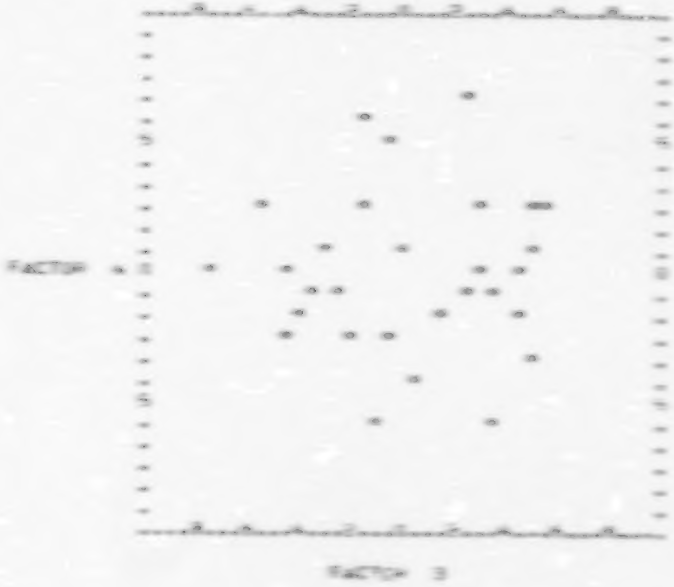
UNROTATED FACTOR LOADINGS



OVERLAP IS INDICATED BY A DOLLAR SIGN. SCALE IS FROM -1 TO +1.

UNROTATED FACTOR LOADINGS





```
OVERLAP IS INDICATED BY A DOLLAR SIGN.  SCALE IS +OR -1 TO +1.

ORTHOGONAL ROTATION, GAMMA =     1.0000

ITERATION    SIMPLICITY
             CRITERION
    0         -4.425436
    1        -10.278742
    2        -11.391173
    3        -11.778440
    4        -11.847086
    5        -11.860543
    6        -11.863196
    7        -11.863573
    8        -11.863645
    9        -11.863667
```

The points in these plots are variables. The coordinates of a variable are its factor loadings. A line or cluster of variables which does not fall on the axes of these plots indicates a potential for simplification of the factors through rotation.

The rotation performed here was orthogonal and required nine iterations using the specified gamma parameter.

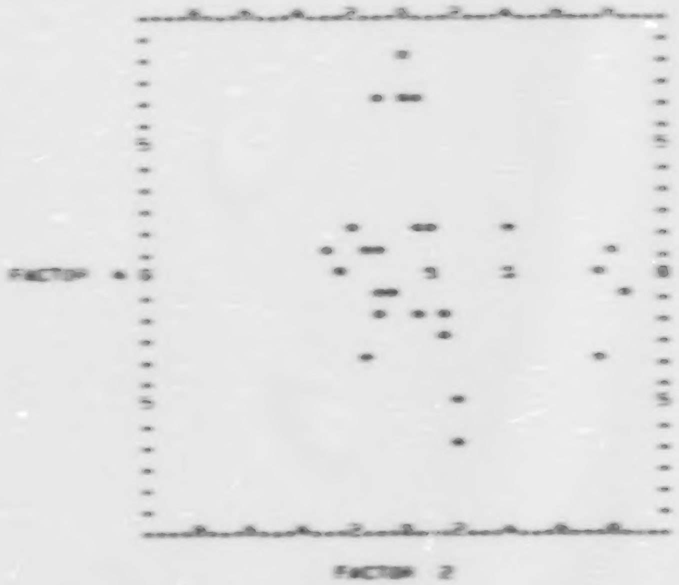
ROTATED FACTOR LOADINGS (PATTERN)

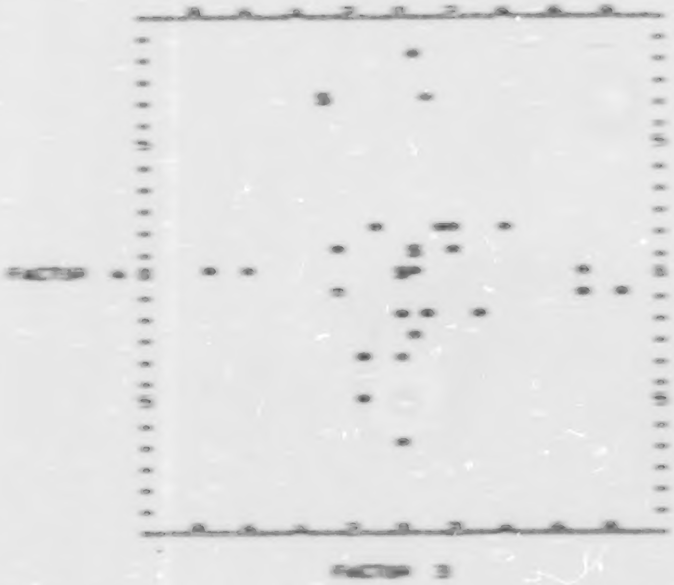
		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
AGE	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FH	7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
RDH	8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PL	9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
WMDH	10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
DDH	11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
FTH	12	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PII	14	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PCS	19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SDHI	21	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HH	22	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	23	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
STH	24	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SLP	25	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
EXT	26	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HC	28	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
SH	29	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
AGTS	30	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TI	31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
TH	32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
II	34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
PHL	41	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HIU	42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ADCI	43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HSU	44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HSBU	45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
EHH	46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
HIGP	47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	VP	4.547	3.394	3.302	3.104	2.457	2.403	2.276

THE VP FOR EACH FACTOR IS THE SUM OF THE SQUARES OF THE ELEMENTS OF THE COLUMN OF THE FACTOR PATTERN MATRIX CORRESPONDING TO THAT FACTOR. WHEN THE ROTATION IS ORTHOGONAL, THE VP IS THE VARIANCE EXPLAINED BY THE FACTOR.

Rotation of the factors has the effect of simplifying the factor loading matrix. Note the rotated matrix has 76 loadings less than 0.2 whereas only 43 of the unrotated factor loadings were less than 0.1. In general, the number of very high and very low loadings is increased at the expense of moderate loadings. Interpretation of the factors is simplified with this procedure.

ROTATED FACTOR LOADINGS





OVERLAP IS INDICATED BY A DOLLAR SIGN. SCALE IS FROM -1 TO +1.

Plots of the rotated loadings are helpful in ascertaining whether the most effective method of rotation has been utilized. If the variables seem to indicate a new axis, an oblique rotation might be tried.

SORTED ROTATED FACTOR LOADINGS (PATTERN)

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
MIGR	44	0.932	0.0	0.0	0.0	0.0	0.0	0.0
NMGR	33	-0.412	0.0	0.0	0.0	0.0	0.0	0.0
PCS	17	0.644	0.0	0.0	0.0	0.417	0.0	0.0
FTW	12	-0.432	0.416	0.391	0.0	0.0	0.0	0.0
PCC	18	0.425	0.0	0.0	0.0	0.0	0.0	0.0
EMT	26	0.0	0.886	-0.251	0.0	0.0	-0.290	0.0
EMM	24	-0.345	0.784	0.0	0.0	0.0	0.0	0.0
EMH	40	0.0	0.754	0.0	0.0	0.0	0.0	0.0
EMF	25	0.0	0.745	0.0	-0.339	0.0	-0.285	0.0
GINI	21	0.385	0.0	0.636	0.0	0.0	0.0	0.0
YM	22	0.0	0.401	-0.744	0.0	0.435	0.0	0.0
NGH	4	-0.281	0.0	0.707	0.0	0.0	0.0	0.0
HC	28	0.0	0.0	0.685	0.0	0.0	0.600	0.0
YF	23	0.0	0.417	-0.597	0.0	0.544	-0.276	0.0
TI	31	0.0	0.0	0.0	0.434	0.0	0.0	0.0
TP	32	-0.409	0.0	0.0	-0.881	0.384	0.0	0.333
MDU	42	0.0	0.0	0.0	0.881	-0.401	0.0	0.0
AUTO	38	0.440	0.0	-0.325	0.867	0.0	0.0	0.0
HOCC	43	-0.334	0.0	-0.304	0.864	0.0	0.0	0.0
HSU	44	0.0	0.0	0.0	-0.514	0.596	0.0	0.0
PU	4	0.0	0.0	0.0	0.0	0.886	0.0	0.0
HSOU	45	0.0	0.0	0.0	-0.294	0.679	0.0	0.0
PM	6	-0.753	0.0	0.0	-0.495	0.629	0.0	-0.447
PHH	11	0.0	0.0	0.0	0.0	0.0	0.482	0.0
AGE	3	-0.442	0.0	0.0	0.0	0.0	-0.776	0.0
FM	7	0.0	-0.294	0.0	0.0	0.0	[illegible]	-0.387
PHL	41	0.0	0.0	0.314	0.0	0.0	0.0	0.786
DI	36	0.0	0.0	0.0	0.0	0.0	0.0	0.739
IM	29	-0.356	0.0	0.0	0.0	-0.272	-0.442	0.580
	VP	4.547	3.384	3.322	3.324	2.432	2.925	2.276

THE ABOVE FACTOR LOADING MATRIX HAS BEEN REARRANGED SO THAT THE COLUMNS APPEAR IN DECREASING ORDER OF VARIANCE EXPLAINED BY FACTORS. THE ROWS HAVE BEEN REARRANGED SO THAT FOR EACH SUCCESSIVE FACTOR, LOADINGS GREATER THAN 0.5000 APPEAR FIRST. LOADINGS LESS THAN 0.2500 HAVE BEEN REPLACED BY ZERO.

This matrix is normally used for interpretation since very low loadings are not printed and high loadings are grouped together. This array has been interpreted as follows:

Factor	Indicates
1	Rapid growth, population heavily young and male, heavy auto dependence
2	High employment, high income, high FTW, low fertility rate
3	Low income—unequal distribution, crowded housing, low auto and home ownership, high FTW and black population
4	High single-family homes, high auto dependence, low female employment, few foreigners
5	Highly urbanized: high income, high density housing, high FTW, low auto dependence
6	Large families, young population, high fertility rate, crowded housing
7	Black, poor health care, few foreigners

FACTOR SCORE COEFFICIENTS
THESE COEFFICIENTS ARE FOR THE STANDARDIZED VARIABLES, MEAN ZERO AND STANDARD DEVIATION ONE.

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
AGE	3	-0.12247	-0.11089	0.12530	0.24474	0.13284	-0.30984	-0.20148
WF	4	-0.10047	-0.11623	0.04044	-0.10446	0.23566	-0.03254	-0.29343
FA	7	-0.05563	0.21067	0.08212	-0.42367	-0.01674	0.29611	-0.11957
MO	8	-0.05771	0.05307	0.23744	0.08448	0.12384	0.02524	0.04040
PU	9	0.03141	-0.03616	0.01366	0.05461	0.29036	0.39251	0.04449
NAOU	10	-0.25746	0.00441	-0.00776	0.02576	-0.00735	0.03444	-0.04710
RHH	11	-0.04565	0.05124	-0.03724	0.02273	0.07371	0.36141	0.07833
FTH	12	-0.13944	0.25766	0.24497	0.16151	0.08243	-0.09462	-0.09421
PCC	20	0.13441	0.05317	0.11447	0.05374	0.09483	-0.03281	-0.05057
PCS	19	0.13873	0.01456	0.03477	0.11474	0.29083	0.08446	0.05047
GINI	21	0.09430	0.07487	0.31245	0.00446	0.01354	-0.07154	-0.02505
TW	22	-0.02244	0.13451	-0.24516	0.14294	0.14442	0.12257	0.09447
TF	23	7.90811	0.02344	-0.11458	0.08677	0.25334	0.08304	0.10162
EMH	24	-0.05863	0.26447	0.03754	0.25444	-0.03813	0.02244	-0.04470
EMF	25	0.09354	0.27443	0.04744	-0.14442	-0.13920	0.01725	-0.07916
EMT	26	0.00845	0.29475	0.04134	-0.03908	-0.04419	0.01843	-0.07514
HC	28	-0.01322	0.15458	0.19993	0.04624	0.10765	0.17451	-0.02314
DH	29	-0.05953	-0.06774	-0.03127	-0.05376	-0.10544	-0.12314	0.22256
AUTO	30	0.07545	0.02623	-0.00475	0.10448	0.03274	0.04692	-0.03232
TU	31	0.00846	0.01694	0.04791	0.33046	0.17348	-0.11640	-0.02435
TP	32	-0.05554	0.03557	-0.06763	-0.19514	0.02773	0.11242	0.25060
DI	38	-0.04445	-0.11044	-0.07691	0.13411	0.14357	-0.01764	0.30526
PHL	41	0.07623	0.11245	0.01571	-0.04354	-0.12674	0.10441	0.35320
MIU	42	-0.00464	0.02238	0.00430	0.21841	-.02264	0.03741	0.13051
MICC	43	-0.11047	0.00538	-0.12251	0.12354	0.04484	0.07444	0.12143
MTL	44	0.04527	-0.00793	-0.03441	-0.10475	0.14681	-0.01597	-0.02425
MSDU	45	0.00508	-0.01025	0.04574	0.05205	0.27367	-0.02636	0.09430
EPM	46	0.02947	0.32083	0.05794	0.02841	-0.01805	0.13425	0.09042
MIGF	48	0.21273	-0.00141	0.03844	-0.04467	-0.05231	-0.09681	-0.04132

This matrix is the inverse of the factor loading pattern (previous page).
Factor score coefficients are used to calculate factor scores:

FACTOR 1 = 0.12 (AGE)-.10 (WF)-.06 (FA)-...

```
FACTOR SCORE COVARIANCE (COMPUTED FROM FACTOR STRUCTURE AND FACTOR SCORE COEFFICIENTS)

                  FACTOR    FACTOR    FACTOR    FACTOR    FACTOR    FACTOR    FACTOR
                     1         2         3         4         5         6         7

  FACTOR   1      1.000
  FACTOR   2     -0.000     1.000
  FACTOR   3      0.000     0.000     1.000
  FACTOR   4      0.000     0.000    -0.000     1.000
  FACTOR   5      0.000    -0.000     0.000    -0.000     1.000
  FACTOR   6      0.000     0.000     0.000     0.000     0.000     1.000
  FACTOR   7      0.000     0.000     0.000    -0.000     0.000    -0.000     1.000

THE DIAGONAL OF THE ABOVE MATRIX CONTAINS THE SQUARED MULTIPLE CORRELATIONS OF EACH FACTOR WITH THE VARIABLES.

  SAVE FILE WRITTEN ON UNIT      [illegible]
                              CODE. . . . IS    SESWDATA
                              CONTENT . IS      DATA
                              LABEL . . IS
                              VARIABLES      1 [illegible]     2 [illegible]     3 AGE       4 [illegible]     5 P
                                             6 [illegible]     7 [illegible]     8 [illegible]     9 [illegible]    10 [illegible]
                                            11 [illegible]    12 [illegible]    13 [illegible]    14 [illegible]    15 [illegible]
                                            16 [illegible]    17 [illegible]    18 [illegible]    19 [illegible]    20 [illegible]
                                            21 [illegible]    22 [illegible]    23 [illegible]    24 [illegible]    25 [illegible]
                                            26 [illegible]    27 [illegible]    28 [illegible]    29 [illegible]    30 [illegible]
                                            31 [illegible]    32 [illegible]    33 [illegible]    34 [illegible]    35 [illegible]
                                            36 [illegible]    37 [illegible]    38 [illegible]    39 LAT       40 LONG
                                            41 [illegible]    42 [illegible]    43 [illegible]    44 [illegible]    45 [illegible]
                                            46 [illegible]    47 [illegible]    48 [illegible]    49 FACTOR1   50 FACTOR2
                                            51 FACTOR3   52 FACTOR4   53 FACTOR5   54 FACTOR6   55 FACTOR7
```

Given an orthogonal rotation, the matrix indicating factor score covariance is of no value. An oblique rotation would result in factors with non-zero covariances which would indicate the interrelationships between these factors.

These statements indicate the location and description of saved output.

```
ESTIMATED FACTOR SCORES AND MAHALANOBIS DISTANCES (CHI-SQUARES) FROM EACH CASE TO THE CENTROID OF ALL CASES
FOR ORIGINAL DATA ( 29 D.F.), FACTOR SCORES ( 7 D.F.) AND THEIR DIFFERENCE ( 22 D.F.).
EACH CHI-SQUARE HAS BEEN DIVIDED BY ITS DEGREES OF FREEDOM.
CASE NUMBERS BELOW REFER TO DATA MATRIX BEFORE DELETION OF MISSING DATA.

      CASE      CHISQ/DF CHISQ/DF CHISQ/DF   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR
  LABEL    NO.        29        7       22        1        2        3        4        5        6        7

ABIL  40    1      0.840    0.883    0.836    0.782   -0.380    0.864    0.271   -1.421   -1.007   -1.276
AKRN  80    2      0.229    0.381    0.181   -0.631    0.826   -0.942    1.651    0.396    0.266    0.227
ALBG 120    3      1.679    0.865    1.927    0.906   -0.333    1.428   -1.361   -0.537    0.000    0.448
ALBN 160    4      1.588    0.425    1.940   -0.640    0.438   -0.917   -1.347   -0.533    0.282   -0.195
ALBQ 200    5      5.218    0.548    6.691    0.417   -1.286    0.521    1.176    0.676   -0.263   -0.221
ALTN 240    7      2.058    0.438    2.319   -0.929    1.883   -0.736   -0.197   -0.710   -0.813   -0.358
ALTO 280    8      0.788    0.826    0.776   -1.702   -0.136    0.438    0.487   -1.081   -0.907    0.250
AMAR 320    9      0.678    0.566    0.734    0.370    0.998    0.295    0.196   -0.865   -0.534   -0.425
ANAH 360   10      1.966    2.257    1.866    1.981    0.217   -0.886    2.373    2.596   -0.865   -0.361
ANDS 400   11      0.698    0.360    0.806   -0.266   -0.281   -1.114    0.664   -0.858   -0.267   -0.172
ANNA 440   12      1.362    1.836    1.472    1.830    1.199   -1.070   -0.091    0.610    0.509   -0.335
APPL 460   13      0.486    0.897    0.629   -0.489    0.480   -0.988    0.441   -0.884    1.450   -0.481
ASHE 480   14      0.959    0.782    0.951   -0.737    0.661    0.876    0.456   -1.295   -1.261    0.577
ATLA 520   15      0.718    0.765    0.783    0.882    1.480   -0.859   -0.382    0.597    0.240    1.240
ATLC 560   16      1.124    1.525    0.967   -1.237   -0.546    1.341   -0.677    0.740   -2.229    1.239
AUGU 600   17      0.529    1.156    0.363    0.338   -0.738    0.076   -0.807   -1.047    0.766    2.281
AUST 640   18      2.562    1.280    2.956    2.431    0.960    0.089   -0.908   -0.021    0.312   -0.256
BAKR 680   19      0.502    0.517    0.510    0.256   -1.440    0.386    1.085   -0.014   -0.261   -0.365
BALT 720   20      0.622    0.792    0.568   -0.639    0.558   -0.237   -1.241    0.180    0.465    1.702
BATN 760   21      0.743    0.343    0.867    0.329    0.386    0.565    0.093    0.334    1.343    1.156
BAYC 800   23      0.658    0.855    0.366   -1.263   -0.757   -1.112    1.134   -0.450    1.026   -0.211
BEAU 840   24      2.296    0.487    2.097   -0.370   -0.118   -0.256    1.332   -0.190    0.676    0.788
BLGS 880   25      0.458    0.498    0.437    0.383   -0.886    0.345    0.356   -0.638   -0.298   -1.033
BLLX 920   26      1.059    1.450    0.933    1.255   -2.382    0.090   -0.740   -1.056    0.677    1.488
BING 960   27      0.744    0.298    0.886   -0.992    0.236   -0.465   -0.146   -0.423   -0.106   -0.881
BIRM1000   28      0.468    0.629    0.416   -0.819    0.083    1.118    0.343   -0.486   -0.329    1.429
BLOO1040   29      1.633    1.878    1.421    0.717    1.525   -0.282   -0.058   -1.656   -0.231   -1.282
BOIS1080   30      0.958    0.851     .836    1.806    0.528    0.268    1.536   -0.293   -0.563   -1.350
BOST1120   31      1.838    1.358    0.980   -0.910    0.611   -0.306   -2.631    1.298    0.212   -0.983
BRDG1160   32      0.481    0.617    0.430   -0.806    0.676   -0.579    0.133    1.950   -0.117   -0.510
BRTS1170   33      0.840    0.672    0.906   -0.693    0.556   -1.417    0.095    1.261    0.367   -0.481
BROK1200   34      0.561    0.309    0.651   -0.560    0.143   -0.828   -0.863    0.225    0.883   -1.086
BRWN1240   35      2.297    0.260    1.868   -1.483   -0.309    0.677    0.236    0.573    3.486   -1.713
BRYN1260   36      2.816    1.286    2.245    2.320    0.236    0.677   -1.236   -0.975    0.813   -0.786
BUFF1280   37      0.425    0.452    0.403   -1.431   -0.492   -0.082   -1.544    0.228   -0.237    0.061
CANT1320   38      0.241    0.390    0.193   -0.961   -0.867   -0.096    0.957   -0.436   -0.188   -0.055
CDRR1360   39      0.283    0.309    0.275    0.122    0.097   -1.057    0.608   -0.348    0.310   -0.335
CHMP1400   40      0.808    1.588    0.667    2.297    0.836   -1.129   -1.762   -1.838    0.860   -0.467
CHRS1440   41      0.853    1.497    0.644    0.613   -1.346    0.282   -1.227   -0.767    1.378    2.044
CHRW1480   42      0.884    0.667    0.863   -1.236   -0.527    0.299   -0.828   -1.305   -1.092    0.463
CHRL1520   43      0.492    0.861    0.396    0.556    1.966    0.383   -0.047   -0.661    0.212    0.472
CHAT1560   44      0.327    0.507    0.270   -0.520    0.066    0.746    0.403   -0.788   -0.825    0.608
CHIC1600   45      0.543    1.451    0.266   -1.707    0.377   -0.638   -2.109    1.066    0.425    1.410
CINC1640   46      0.546    0.153    0.671   -0.600    0.108   -0.272   -0.082    0.075    0.238    0.421
CLEV1680   47      0.357    0.527    0.302   -0.878    0.440   -0.065   -0.489    1.280   -0.833    0.734
CPLD1720   48      1.413    2.452    0.923    2.406   -2.938   -1.057    0.116    1.254   -0.282    0.572
COLM1740   49      1.357    2.280    1.061    2.681    1.938    0.228   -1.258   -1.275   -0.795   -1.323
COLS1760   50      0.519    0.603    0.429    0.905   -0.832    0.077   -0.744   -1.805    1.045    1.457
COLU1800   51      0.904    1.423    0.739    1.351   -1.447    0.561   -1.396   -0.729    0.542    1.786
COLU1840   52      0.266    0.255    0.262    0.451    1.809   -0.512   -0.466    0.880    0.146    0.387
```

This table describes each SMSA
in terms of the seven factors
selected. Chi-Square values
quantify the eccentricity of each
observation. The factor scores
locate each observation with
respect to each axis. These are
the results which are input to
the cluster analysis which
follows.

ESTIMATED FACTOR SCORES AND MAHALANOBIS DISTANCES (CHI-SQUARES) FROM EACH CASE TO THE CENTROID OF ALL CASES
FOR ORIGINAL DATA (29 D.F.) FACTOR SCORES (7 D.F.) AND THEIR DIFFERENCE (22 D.F.).
EACH CHI-SQUARE HAS BEEN DIVIDED BY ITS DEGREES OF FREEDOM.
CASE NUMBERS BELOW REFER TO DATA MATRIX BEFORE DELETION OF MISSING DATA.

CASE LABEL	NO.	CHISQ/DF 29	CHISQ/DF 7	CHISQ/DF 22	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
CLMP1880	53	1.025	1.448	0.523	-1.136	-1.026	1.583	0.736	0.436	1.830	-0.288
DALS1920	54	0.488	1.804	0.392	1.045	1.021	0.278	0.336	1.476	0.888	0.742
DAVN1740	56	0.375	0.338	0.384	-0.504	-0.387	-1.119	0.678	0.254	-0.873	0.672
DAYT2000	57	0.292	0.323	0.243	-0.053	0.593	-1.145	0.501	-0.888	0.356	0.459
DCAT2040	59	1.437	1.486	1.041	-0.456	0.207	-0.627	1.237	0.138	-0.323	1.046
DENV2080	60	0.344	0.357	0.345	0.431	0.493	-0.463	0.224	0.898	-0.361	-0.492
DESM2120	61	0.365	0.290	0.383	-0.128	1.266	-0.454	0.248	-0.047	-0.211	-0.336
DETR2160	62	0.577	0.988	0.454	-0.477	-0.028	-0.963	0.748	1.556	0.864	1.674
DUMU2200	63	2.183	1.923	2.290	-0.818	1.175	-0.765	0.888	-0.189	3.212	2.176
DULU2240	64	0.637	0.913	0.564	-1.267	-1.567	-0.549	-0.759	-1.738	-0.375	-0.885
DHAM2280	65	1.025	1.384	1.388	1.251	1.775	0.421	-1.318	-1.035	-0.854	0.527
FLPS2320	66	1.118	2.012	0.761	-0.892	-1.813	1.388	-0.702	1.492	2.183	-1.289
EVIE2360	68	1.377	1.407	0.384	-1.233	-0.484	-0.515	0.298	-0.763	-0.852	-0.452
EUGN2400	69	1.028	1.678	0.461	1.391	-0.326	-0.367	0.782	-0.483	-0.468	-1.484
FARG2440	70	0.494	1.383	0.481	-0.891	0.129	1.324	0.739	-0.677	-0.731	-0.298
FALW2480	71	1.196	1.831	1.085	-1.225	-0.456	1.441	-1.078	0.227	-1.565	-2.133
FARG2520	72	1.294	1.421	1.198	0.842	1.619	-0.331	-0.419	-0.596	1.115	-1.256
FAYT2560	73	2.118	3.945	1.518	2.405	-1.567	-0.948	-1.542	-0.867	1.384	1.456
FLNT2640	75	0.413	0.992	0.224	-0.615	-0.254	-0.341	1.382	0.876	1.457	0.654
FTLA2680	77	2.657	5.883	1.408	1.644	0.021	1.408	2.454	3.408	-0.719	-0.848
FTSM2720	78	0.733	1.325	0.640	-0.271	-0.604	1.596	0.625	-1.318	-1.288	-0.384
FTWA2760	80	0.414	0.532	0.635	-0.234	0.481	-1.249	0.787	-0.375	0.781	0.178
FTWO2800	81	0.573	0.548	0.511	0.724	0.774	-0.286	1.087	0.788	0.888	0.485
FRES2840	82	1.040	1.193	1.415	0.307	-1.109	1.324	-0.544	-0.136	-0.460	-1.794
GAYS2880	83	0.436	1.294	0.847	-1.056	-1.052	0.819	1.571	-0.881	-0.292	1.138
GRIS2920	84	3.395	4.894	3.173	2.421	0.653	2.873	-2.831	-1.653	-0.462	-1.483
GRLA2960	85	3.296	0.871	3.967	0.766	0.286	0.812	-2.110	-0.466	-0.486	-0.777
GRNT3000	86	0.726	1.874	0.621	0.986	-0.263	-1.121	0.293	1.866	1.865	1.456
GRAP3080	88	0.444	0.508	0.431	-0.576	-0.887	-1.112	0.881	-0.225	0.848	-0.603
GTFA3040	89	1.142	1.574	1.030	0.944	-2.477	-0.688	-0.488	-0.429	-0.535	-0.784
GREE3120	90	1.874	1.153	2.109	-0.254	-1.448	-1.143	0.976	0.957	1.648	-0.234
HARS3120	91	0.844	0.888	0.875	-0.215	1.727	0.443	0.411	-0.856	-0.251	1.428
HARR3160	92	1.047	0.732	1.614	-0.254	1.357	0.393	0.345	-1.388	-0.816	1.052
HARL3200	93	0.834	1.204	0.717	-0.132	0.172	-0.588	0.736	-0.438	0.489	-0.294
HATS3240	94	0.331	0.427	0.311	-0.796	0.784	-0.358	-1.125	-0.876	-0.762	0.234
HART3280	95	0.545	1.639	0.525	-0.478	1.708	-1.104	-0.657	1.074	0.128	-0.433
HOUS3360	96	0.740	0.703	0.742	0.623	0.425	0.364	0.621	1.346	0.578	1.121
HUTN3400	97	0.981	0.848	0.973	-1.088	-0.486	0.816	0.348	-1.336	-1.184	-0.231
HNTS3440	98	1.082	1.531	1.125	1.124	0.388	0.544	1.217	0.143	0.889	0.313
INDI3480	99	0.446	0.303	0.486	0.874	0.746	-0.583	0.254	0.262	0.266	0.485
JMOC3520	100	1.176	0.846	1.288	-0.614	-1.127	-1.498	0.978	-0.785	0.175	0.626
JMSN3560	101	0.636	1.138	0.734	-0.888	0.702	1.586	0.497	-0.346	0.888	1.865
JFLA3600	102	1.424	0.555	2.388	1.844	0.238	1.025	0.896	0.322	-0.368	1.288
JEMS3640	103	2.518	4.419	1.758	-1.288	-0.888	0.388	-5.379	1.536	-1.254	0.231
JOHN3680	104	0.802	1.356	0.627	-2.110	-1.496	0.824	-0.243	-1.424	-0.768	-0.875
KANS3720	105	1.203	0.217	1.296	0.102	0.965	-0.438	0.101	0.744	-0.255	0.442
KENO3800	107	0.556	0.476	0.601	-0.638	-0.227	-1.611	0.106	-0.173	0.788	-0.362
KNOX3840	108	0.868	0.463	0.988	-0.144	0.272	1.328	0.675	-0.544	-0.988	0.888
LFLA3880	111	1.546	1.383	1.659	0.722	0.449	1.569	0.298	-0.848	2.882	-0.246
LFIN3920	112	0.925	1.311	0.603	1.957	1.282	-0.914	-0.888	-1.388	0.461	-1.267

ESTIMATED FACTOR SCORES AND MAHALANOBIS DISTANCES (CHI-SQUARES) FROM EACH CASE TO THE CENTROID OF ALL CASES
FOR ORIGINAL DATA (29 D.F.), FACTOR SCORES (7 D.F.) AND THEIR DIFFERENCE (22 D.F.).
EACH CHI-SQUARE HAS BEEN DIVIDED BY ITS DEGREES OF FREEDOM.
CASE NUMBERS BELOW REFER TO DATA MATRIX BEFORE DELETION OF MISSING DATA.

CASE LABEL	NO.	CHISQ/DF 29	CHISQ/DF 7	CHISQ/DF 22	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
LAMC3460	113	0.447	0.585	0.675	-0.606	-0.046	0.442	0.670	-0.855	1.260	0.565
LAMC4080	115	0.800	0.773	0.935	-0.680	1.143	-0.861	-0.198	-1.882	0.827	-0.796
LAMS4040	116	0.440	0.640	0.376	0.711	0.837	-1.279	0.040	-0.405	1.040	-0.848
LARE4080	117	2.608	0.508	1.883	-1.606	-2.483	4.467	-0.198	1.406	3.385	-1.671
LASA4120	118	2.609	2.277	2.717	2.471	0.216	-4.247	1.485	2.492	-0.312	1.087
LAWA4160	119	0.930	0.816	0.983	-0.745	0.711	-0.235	-0.985	0.740	-0.487	-1.754
LAWA4200	120	2.429	3.944	1.930	2.377	-4.072	-1.310	-1.481	-0.587	0.705	1.081
LEWS4240	121	1.087	1.380	1.067	-0.426	-0.563	0.670	-2.234	-0.833	-0.963	-1.561
LEXA4280	122	0.713	0.600	0.762	1.480	1.090	0.585	-0.354	0.411	-0.546	-0.896
LIMA4320	123	3.530	0.492	4.507	-3.181	-6.347	-0.961	0.545	-1.380	0.376	0.180
LIMC4360	124	0.725	1.376	0.510	1.047	1.544	-0.384	-1.001	-1.390	-0.256	-1.456
LITL4400	125	0.274	0.380	0.241	0.391	0.147	0.743	0.391	-0.406	-0.722	0.537
LOPA4440	127	1.029	0.756	1.110	-0.343	-0.831	-0.743	1.516	0.776	1.203	0.463
LTSA4480	128	1.500	0.675	1.773	0.391	0.237	0.130	-0.296	1.650	-1.203	-0.524
LPUS4520	129	2.900	0.480	3.776	-0.247	-1.261	-0.325	-0.120	0.342	0.280	1.218
LWEL4560	130	0.705	0.821	0.732	-0.736	-0.276	-0.911	-0.443	0.946	1.167	-0.764
LUPA4602	131	0.800	0.438	0.757	1.14	0.267	0.796	0.752	-0.070	0.558	-0.291
LYMC4640	132	0.900	0.433	0.847	-0.430	1.265	0.932	-0.120	-1.096	-0.208	1.590
MACH4680	133	1.095	0.717	0.890	0.507	0.813	1.209	-0.519	-0.556	0.487	1.485
MADA4720	134	0.848	1.295	0.445	1.136	1.116	-0.916	-1.637	-0.386	0.880	-1.532
MARC4760	135	0.562	0.940	0.441	-1.086	-0.296	-0.113	-1.965	0.230	-0.588	-1.596
MARS4800	136	0.332	0.384	0.321	-0.375	0.817	-1.081	0.713	-0.862	-0.831	0.428
MCAL4880	137	2.593	7.070	0.880	-0.650	-0.382	5.031	0.581	0.857	4.781	-1.480
MEMP4920	138	0.572	1.003	0.435	1.254	0.476	1.334	-0.858	0.046	0.865	1.675
MESC4960	140	0.857	0.550	0.845	-0.958	0.369	-0.560	-0.192	1.597	-0.123	-0.690
MIES5000	141	1.737	3.230	1.289	0.135	0.782	2.744	0.077	3.442	-0.588	0.071
MIDL5040	142	1.394	0.757	1.863	0.546	1.423	0.321	1.501	0.897	2.457	-0.585
MILW5080	143	0.303	0.480	0.290	-0.695	0.451	-1.088	-0.982	0.493	0.516	-0.323
MINN5120	144	0.420	0.737	0.320	-0.127	1.612	-1.251	-0.556	0.924	1.032	-0.567
MOBL5160	145	0.443	0.840	0.410	-0.576	-1.369	1.125	0.290	-0.630	0.916	1.234
MODE5170	146	0.853	1.045	0.529	0.411	-1.306	0.481	1.294	-0.121	-1.071	-1.301
MONT5200	147	0.895	0.734	0.880	-0.240	-0.106	1.724	0.605	-0.921	0.173	0.855
MONT5240	148	0.837	1.391	0.677	-0.257	0.365	1.465	-0.253	-0.830	0.813	2.520
MORC5280	149	1.882	0.560	1.427	-0.136	-1.813	-0.865	1.146	-0.385	-0.311	0.646
MSPR5320	150	0.473	0.805	0.425	-1.056	-0.470	-0.806	1.098	-0.393	1.042	0.395
NALL5360	152	1.724	0.710	1.858	0.860	1.390	1.213	0.274	0.341	-0.518	0.610
NBED5400	153	1.209	1.635	1.073	-1.343	-0.255	0.410	-1.488	0.223	-1.080	-1.823
NBPT5440	154	0.762	0.806	0.780	-0.713	0.808	-0.647	-0.456	1.146	-0.234	-1.276
NHAV5480	156	0.329	0.483	0.304	-0.844	0.750	-2.188	-0.886	0.675	-0.572	-0.240
NLIN5520	157	0.863	0.537	0.940	0.146	-1.105	-0.407	-0.549	-0.386	0.058	-0.274
NORL5560	158	0.984	1.304	0.885	-0.517	-0.898	1.278	-1.881	0.254	0.780	1.780
NPDR5600	159	4.382	4.041	3.656	-1.406	0.216	0.626	-4.718	3.630	-0.632	1.592
NPDR5640	160	0.508	1.143	0.405	-0.773	0.829	-0.731	-1.194	2.250	-0.326	0.475
NPOR5680	161	0.720	1.267	0.570	1.173	-0.801	-0.953	-0.928	0.306	1.056	1.603
NURF5720	162	0.845	1.636	0.593	1.257	-1.882	-0.574	-1.498	-0.102	0.672	1.087
NOHM5760	163	3.516	1.853	4.380	-0.142	1.708	-0.481	0.414	1.837	0.021	0.354
ODES5800	164	1.414	0.715	1.636	0.230	0.382	-0.103	2.006	0.250	0.548	0.045
OGDE5840	165	0.813	0.735	0.576	-0.862	0.361	-0.719	0.237	-0.381	1.583	-1.394
OKLA5880	166	0.514	0.344	0.568	0.788	0.762	0.053	0.854	0.103	-0.493	-0.860
OMAH5921	167	0.257	0.175	0.293	0.142	0.352	-0.591	-0.486	0.356	0.672	-0.230

ESTIMATED FACTOR SCORES AND MAHALANOBIS DISTANCES (CHI-SQUARES) FROM EACH CASE TO THE CENTROID OF ALL CASES
FOR ORIGINAL DATA (29 D.F.) FACTOR SCORES (7 D.F.) AND THEIR DIFFERENCE (22 D.F.).
EACH CHI-SQUARE HAS BEEN DIVIDED BY ITS DEGREES OF FREEDOM.
CASE NUMBERS BELOW REFER TO DATA MATRIX BEFORE DELETION OF MISSING DATA.

CASE LABEL	NO.	CHISQ/DF 29	CHISQ/DF 7	CHISQ/DF 22	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
OKLA5880	169	1.089	0.433	0.608	0.751	0.840	0.525	1.032	0.439	-0.431	0.480
PATE6040	172	1.405	1.446	1.025	-1.007√	0.491	-0.514	-0.043	2.003√	-0.346	-0.338
PENS6080	173	1.465	0.638	0.333	0.474	-1.317	0.025	0.290	-0.557	0.596	1.384
PEOR6120	174	0.453	0.325	0.494	-0.542	0.117	-1.182	0.614	-0.261	0.614	0.387
PHIL6160	176	0.733	0.841	0.693	-1.258	0.243	-0.434	-0.865	0.629	0.254	1.635
PHOE6200	177	0.363	1.047	0.257	0.496	-0.021	0.429	1.371√	1.274√	-0.128	-0.541
PITT6280	179	0.333	0.762	0.468	-1.845	-0.457	-0.331	-0.305	0.354	-0.704	0.480
PITT6320	180	0.317	0.629	0.282	-1.211	-0.324	-0.933	-0.201	0.013	-0.383	-0.625
PORT6400	181	0.435	0.541	0.667	-0.960	-0.354	-0.046	-1.007	-0.198	-0.973	-0.673
PORT6440	182	0.344	0.259	0.373	0.043	0.104	-0.548	0.329	0.438	-0.862	-0.395
PROV6480	184	0.656	0.329	0.762	-0.931	-0.041	-0.223	-0.654	0.438	-0.598	-0.544
PROV6520	185	1.355	1.629	1.174	0.754	-0.436	0.123	0.086	-0.472	2.719	-2.233
PUEB6560	186	1.404	0.541	0.637	-0.404	-0.134	-0.803	1.255	0.174	0.432	4.394
RACI6600	187	0.302	0.652	0.252	-0.493	0.836	-1.145	0.320	0.887	1.138	-1.524
RALE6640	188	0.499	0.882	0.445	1.154	1.423	0.298	-0.301	-0.750	0.252	0.725
READ6680	189	0.889	0.814	0.676	-1.430	0.404	-0.333	0.252	-0.888	-1.375	-1.053
RENO6720	190	1.194	1.084	1.225	1.511	1.520	-0.182	0.387	0.627	-1.354	-0.813
RICH6760	192	0.633	0.447	0.519	-0.105	1.506	0.151	-0.729	-0.896	-0.221	1.888
RIVE6780	193	1.624	0.862	1.005	0.094	-1.561	0.445	0.747	0.598	-0.837	-1.494
ROAN6800	194	0.844	0.397	0.796	-0.081	1.143	0.653	0.396	-0.241	-0.641	0.159
ROCH6820	195	1.014	1.257	1.013	0.575	1.365	-0.629	-0.219	-0.504	0.838	-1.492
ROCH6840	196	0.329	0.350	0.321	-0.671	0.434	-1.148	-0.227	0.444	0.335	-0.096
ROCK6880	197	0.253	0.377	0.251	-0.081	0.707	-1.017	0.310	0.090	0.450	-0.582
SACR6920	198	0.314	0.395	0.290	0.633	-0.762	-0.378	0.725	0.766	-0.462	-0.253
SAGI6960	199	0.414	1.156	0.211	-0.449	-0.605	-1.264	1.225	0.081	1.546	0.744
SJOS7000	200	0.915	0.727	0.975	-0.891	-0.761	0.241	0.335	-0.538	-1.666	0.387
STLO7040	201	0.480	0.227	0.561	-0.765	0.144	-0.346	-0.201	0.221	0.195	0.867
SALE7080	202	1.245	1.404	1.167	0.521	-1.921	0.017	0.546	-0.540	-1.532	-1.377
SALI7120	203	1.247	1.329	1.244	1.097	-2.005	-0.559	-0.617	0.217	0.208	-0.549
SALT7160	204	0.805	0.637	0.464	-0.112	-1.564	-0.701	0.447	0.611	1.748	-1.175
SANA7200	205	0.592	0.677	0.559	0.721	-0.415	1.490	0.631	-0.775	-0.734	-0.632
SANT7240	206	0.825	0.754	0.582	0.898	-1.094	1.228	0.180	1.082	1.360	0.244
SDIE7320	207	0.650	1.149	0.502	1.579	-1.457	-0.893	-0.470	0.882	-0.461	-0.244
SFRA7360	208	0.696	0.817	0.458	0.145	0.211	-0.752	-1.258	1.571	-1.035	0.873
SJOS7400	209	0.859	1.720	0.385	1.398	0.391	-0.774	1.494	2.439	0.459	-0.446
SBAR7480	210	0.631	1.024	1.544	1.752	-0.462	-0.258	0.304	1.204	-0.712	-1.156
SROS7500	212	1.307	1.294	1.320	0.696	-1.546	0.383	1.241	-0.010	-1.456	-0.936
SAVA7520	214	0.858	0.471	0.823	-0.274	-0.331	0.732	-0.955	-1.085	-0.938	2.024
SCRA7560	215	1.312	1.429	1.276	-2.087	-0.554	0.428	-0.852	-0.293	-1.077	-0.598
SEAT7600	216	0.533	1.456	0.559	0.356	0.856	-1.272	0.873	1.880	-0.484	-0.298
SHER7640	217	0.857	1.042	0.767	0.492	0.329	1.185	0.134	-1.679	-1.418	-0.848
SHRE7680	218	0.463	1.632	0.194	-0.171	-0.250	1.319	0.398	-0.603	0.070	1.454
SIOU7720	219	0.427	1.175	0.508	-0.661	0.052	0.141	0.298	-0.642	-0.147	-0.554
SIOU7760	221	0.493	1.092	0.838	0.074	1.437	0.143	-0.893	-1.346	0.888	-1.847
SBEN7800	222	0.212	0.310	0.178	-0.845	0.214	-0.676	0.632	-0.525	0.357	-0.810
SPOK7840	223	0.456	0.741	0.569	-0.093	-0.630	-0.328	0.257	-0.229	-0.853	-0.457
SPRI7880	224	1.062	0.406	1.251	-0.485	1.244	-0.178	-0.156	-0.404	-0.687	-0.956
SPRI7920	225	0.677	0.637	0.881	0.527	0.360	0.515	0.412	-0.783	-1.525	-0.066
SPRI8000	226	0.432	0.220	0.580	-0.202	0.234	-0.523	0.664	-0.634	-0.761	-0.150
STAM8040	227	2.370	0.698	2.905	-0.747	-0.405	-0.724	-1.654	0.211	-0.287	-0.835

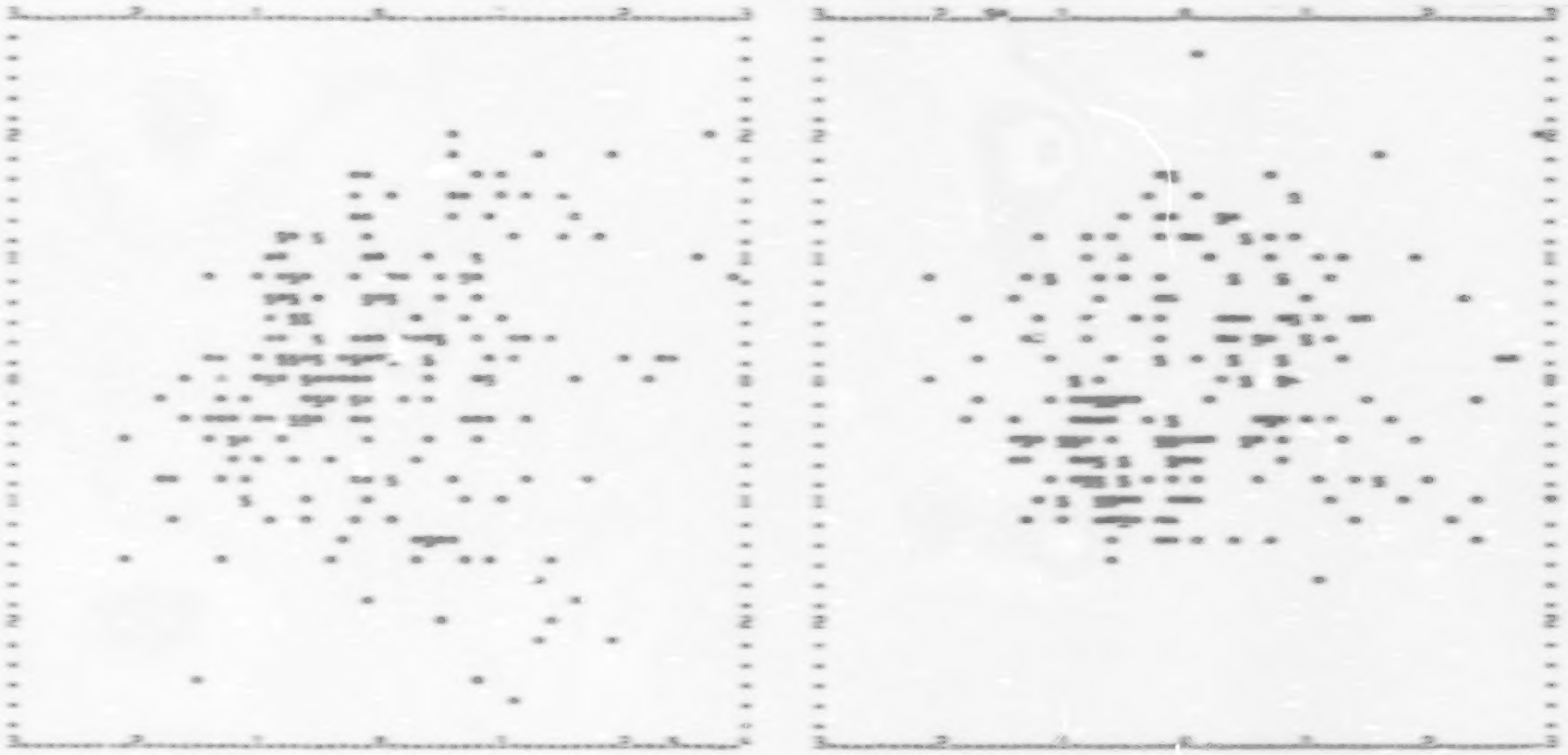
```
ESTIMATED FACTOR SCORES AND MAHALANOBIS DISTANCES (CHI-SQUARES) FROM EACH CASE TO THE CENTROID OF ALL CASES
FOR ORIGINAL DATA ( 29 D.F.) FACTOR SCORES ( 7 D.F.) AND THEIR DIFFERENCE ( 22 D.F.).
EACH CHI-SQUARE HAS BEEN DIVIDED BY ITS DEGREES OF FREEDOM.
CASE NUMBERS BELOW REFER TO DATA MATRIX BEFORE DELETION OF MISSING DATA.

      CASE     CHISQ/DF CHISQ/DF CHISQ/DF   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR   FACTOR
  LABEL    NO.       29        7       22        1        2        3        4        5        6        7

STEWARRO 229     1.266    0.326    1.386   -1.651   -1.219   -0.243    0.895    0.407   -0.714    0.352
STOCK120 230     1.082    1.431    1.303   -0.238   -1.185    0.892    0.453    0.553   -0.360   -0.061
SYRAM140 231     0.829    0.252    1.025   -0.533   -0.211   -0.692   -0.536   -0.475    0.807   -0.413
TACOM200 232     0.961    1.623    0.750    1.076   -2.058   -1.081   -0.547    0.185   -0.147    0.287
TALLB240 233     1.276    1.376    1.245    1.947    1.906    0.423   -0.513   -0.968    0.135    0.356
TAMPA240 234     1.415    1.722    1.366    0.183   -0.893    1.482    1.309    1.476   -2.274    0.123
TEXAR320 235     1.246    0.968    1.334   -0.727   -0.647    0.656    0.281   -1.385   -1.344    0.282
TEXAR360 236     1.250    0.667    1.449   -0.676   -0.997    1.628    0.396   -1.272   -0.865    0.960
TOLEP400 237     0.818    0.269    0.987   -0.678    0.171   -0.773    0.692   -0.296    0.636    0.256
TOPE3440 238     0.882    0.196    0.638    0.456    0.182   -0.496    0.129   -0.629   -0.856   -0.668
TREN4480 239     0.856    0.440    0.786   -0.679    0.765   -0.277   -0.611    0.639   -0.552    1.120
TUSC2520 240     0.773    0.738    0.796    0.742   -0.923    0.427    0.910    1.056   -0.385   -0.413
TULS2560 241     0.499    0.260    0.580    0.354    0.270    0.288    0.629   -0.094   -0.477   -0.182
TUSCA602 242     1.567    0.505    1.886    0.416   -0.456    1.388    0.136   -1.095   -0.644    0.141
TYLH4640 243     0.899    0.529    1.016    0.225    0.876    1.026    0.632   -0.628   -0.629    0.592
UTIC6680 244     2.242    0.432    2.419   -0.644   -0.932   -0.563   -0.806   -0.495   -0.596   -0.412
VALLA720 245     1.366    1.129    1.436    1.356   -2.076   -1.223    0.055    0.460   -0.183    0.133
VINEE760 246     0.793    0.256    1.955   -0.882    0.590    0.665    0.256   -0.436   -0.671   -0.298
WACO8800 247     0.899    0.489    1.055   -0.243    0.819    0.783    0.769   -0.554   -0.798    0.857
WASH8840 248     2.183    1.068    2.541    0.627    1.767   -0.868   -0.566    0.471    0.420    1.592
WATT8880 249     2.520    1.119    2.976   -0.786    0.237   -0.093   -0.623    2.535   -0.425   -0.520
WYTO8920 250     0.425    0.385    0.458   -0.378    0.166   -0.547    0.547   -0.618    0.758   -0.656
WPAL8960 251     2.025    1.095    2.013    0.987    0.187    1.486    1.196    2.343   -1.898    0.677
WEEE9000 252     0.631    1.043    0.580   -1.727   -0.368    0.104   -0.376   -1.294   -1.308   -0.272
WICH9040 253     0.525    0.151    0.644    0.196   -0.171   -0.542    0.632   -0.431   -0.327   -0.656
WIFA9080 256     2.111    0.589    2.610    1.837   -1.819    0.632    0.814   -0.730   -0.753    0.156
WCJ9140  257     0.525    0.215    0.624   -0.677    0.291   -0.765    0.120    0.888    0.442    0.659
WILM9200 258     0.914    2.072    0.936    0.828    0.250    1.208   -0.488   -1.596   -0.637    1.223
WORC9240 259     0.377    0.594    0.338   -1.262    0.866   -0.744   -0.798    0.290   -0.211   -0.598
YORK9280 261     0.429    0.776    0.860   -0.868    0.763   -0.556    0.437   -1.637   -0.450   -0.390
YOUN9320 262     0.447    0.624    0.340   -1.431   -0.344   -0.463    1.796    0.232   -0.831    0.266

SAVE FILE ON UNIT  9 HAS BEEN COMPLETED.
```

FACTOR SCORE COVARIANCE (COMPUTED FROM FACTOR SCORES)

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
FACTOR 1	1.000						
FACTOR 2	-0.000	1.000					
FACTOR 3	-0.000	-0.000	1.000				
FACTOR 4	0.000	0.000	0.000	1.000			
FACTOR 5	0.000	0.000	0.000	-0.000	1.000		
FACTOR 6	-0.000	-0.000	0.000	-0.000	0.000	1.000	
FACTOR 7	-0.000	0.000	-0.000	-0.000	0.000	-0.000	1.000

Again, this array is of no value, since the rotation has constrained the factors to have zero covariance.

FACTOR SCORES



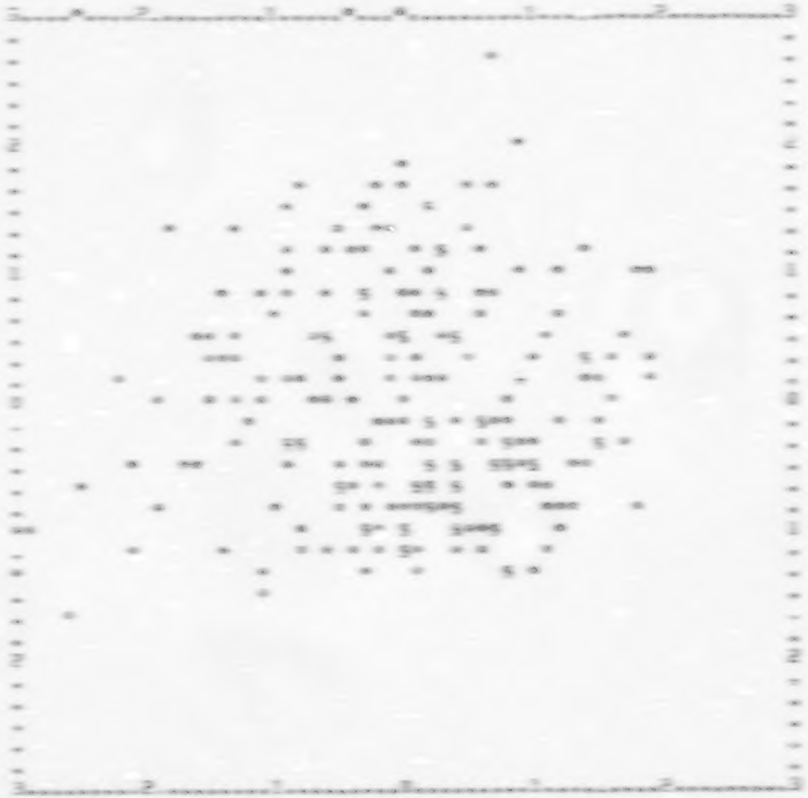
X-AXIS IS FACTOR 1, Y-AXIS IS FACTOR 2

X-AXIS IS FACTOR 1, Y-AXIS IS FACTOR 3

OVERLAP IS INDICATED BY A DOLLAR SIGN. SCALE IS FROM -3 TO +3.
FACTOR SCORES GREATER THAN 3 ARE PLOTTED AS 3. FACTOR SCORES LESS THAN -3 ARE PLOTTED AS -3.

FACTOR SCORES

X-AXIS IS FACTOR 2, Y-AXIS IS FACTOR 3



OVERLAP IS INDICATED BY A DOLLAR SIGN. SCALE IS FROM -3 TO +3.
FACTOR SCORES GREATER THAN 3 ARE PLOTTED AS 3. FACTOR SCORES LESS THAN -3 ARE PLOTTED AS -3.

These plots are helpful in spotting outliers. Where only two or three factors are being utilized, these plots can be used to identify groups of observations.

```
BMDP4M - FACTOR ANALYSIS - DOUBLE PRECISION VERSION
HEALTH SCIENCES COMPUTING FACILITY                        PROGRAM REVISED JULY 7, 1975
UNIVERSITY OF CALIFORNIA, LOS ANGELES                     MANUAL DATE  --  1975

 PROGRAM CONTROL INFORMATION

           FINISH/

 PROGRAM TERMINATED NORMALLY.
```

```
End of factor analysis program.
```

```
BMDP2M - CLUSTER ANALYSIS OF CASES                       PROGRAM REVISED JULY 7, 1975
HEALTH SCIENCES COMPUTING FACILITY                       MANUAL DATE -- 1975
UNIVERSITY OF CALIFORNIA, LOS ANGELES


PROGRAM CONTROL INFORMATION

          /PROBLEM     TITLE IS 'FIRST CLUSTERING FOR SCS'.
          /INPUT UNIT IS 4.  CONTENT IS DATA. CODE IS SCSDATA.
          /VARIABLES  USE ARE 49 TO 55.  LABEL IS 1,2.
            /PRINT YES.
            / END / FINISH /


PROBLEM TITLE . . . . . . . . FIRST CLUSTERING FOR SCS

NUMBER OF VARIABLES TO READ IN. . . . . . . . . . .      55
NUMBER OF VARIABLES ADDED BY TRANSFORMATIONS. . .         0
TOTAL NUMBER OF VARIABLES . . . . . . . . . . . . .      55
NUMBER OF CASES TO READ IN. . . . . . . . . . . . . INFINITE
CASE LABELING VARIABLES . . . . . . . . . . . . . .    RID1        RID2
LIMITS AND MISSING VALUE CHECKED BEFORE TRANSFORMATIONS
ALWAYS ARE. . . . . . . . . . . . . . . . . . . . . . ZEROS
INPUT UNIT NUMBER . . . . . . . . . . . . . . . . .       4
REWIND INPUT UNIT PRIOR TO READING. . DATA. . .         YES

INPUT FORMAT
       SAVE FILE . . . . . . CODE. . . . IS      SCSDATA
                             CONTENT . IS      DATA
                             LABEL . . . IS
                             VARIABLES     1  RID1       2  RID2        3  AGE         4  PW          5  P
                                           6  PE         7  FA          8  SEX         9  PU         10  NMOV
                                          11  [illegible] 12  FTR      13  MS         14  HSMA       15  HS
                                          16  HSG       17  HTOT       18  PCC        19  PCS        20  YMC
                                          21  SSNC      22  TW         23  YP         24  EPA        25  EMF
                                          26  EPT       27  EBL        28  HC         29  DR         30  AUTO
                                          31  TO        32  TP         33  TWR        34  TJA        35  TJU
                                          36  II        37  WAT        38  HOG        39  LAT        40  LONG
                                          41  PAL       42  HSU        43  HOCC       44  HSU        45  HSBU
                                          46  EPB       47  WATP       48  HSWP       49  FACTOR1    50  FACTOR2
                                          51  FACTOR3   52  FACTOR4    53  FACTOR5    54  FACTOR6    55  FACTOR7

VARIABLES TO BE USED
          49  FACTOR1      50  FACTOR2      51  FACTOR3      52  FACTOR4      53  FACTOR5
          54  FACTOR6      55  FACTOR7

NUMBER OF CASES READ. . . . . . . . . . . . . . . .     262
   CASES WITH DATA MISSING OR BEYOND LIMITS . .          71
      REMAINING NUMBER OF CASES . . . . . . . . .       191
PRINT DISTANCE MATRIX . . . . . . . . . . . . . .        NO
TYPE OF TREE PRINTED. . . . . . . . . . . . . . . . REGULAR
CALCULATING PROCEDURE . . . . . . . . . . . . . . . SUM-SUM
STANDARDIZATION OF INPUT DATA . . . . . . . . . .       YES
```

Input to clustering program. Indicates location of saved factor scores.

Interpretation of instructions concerning data. Clustering algorithm will be applied to factor scores on the seven factors.

Clustering method to be used and output to be printed.

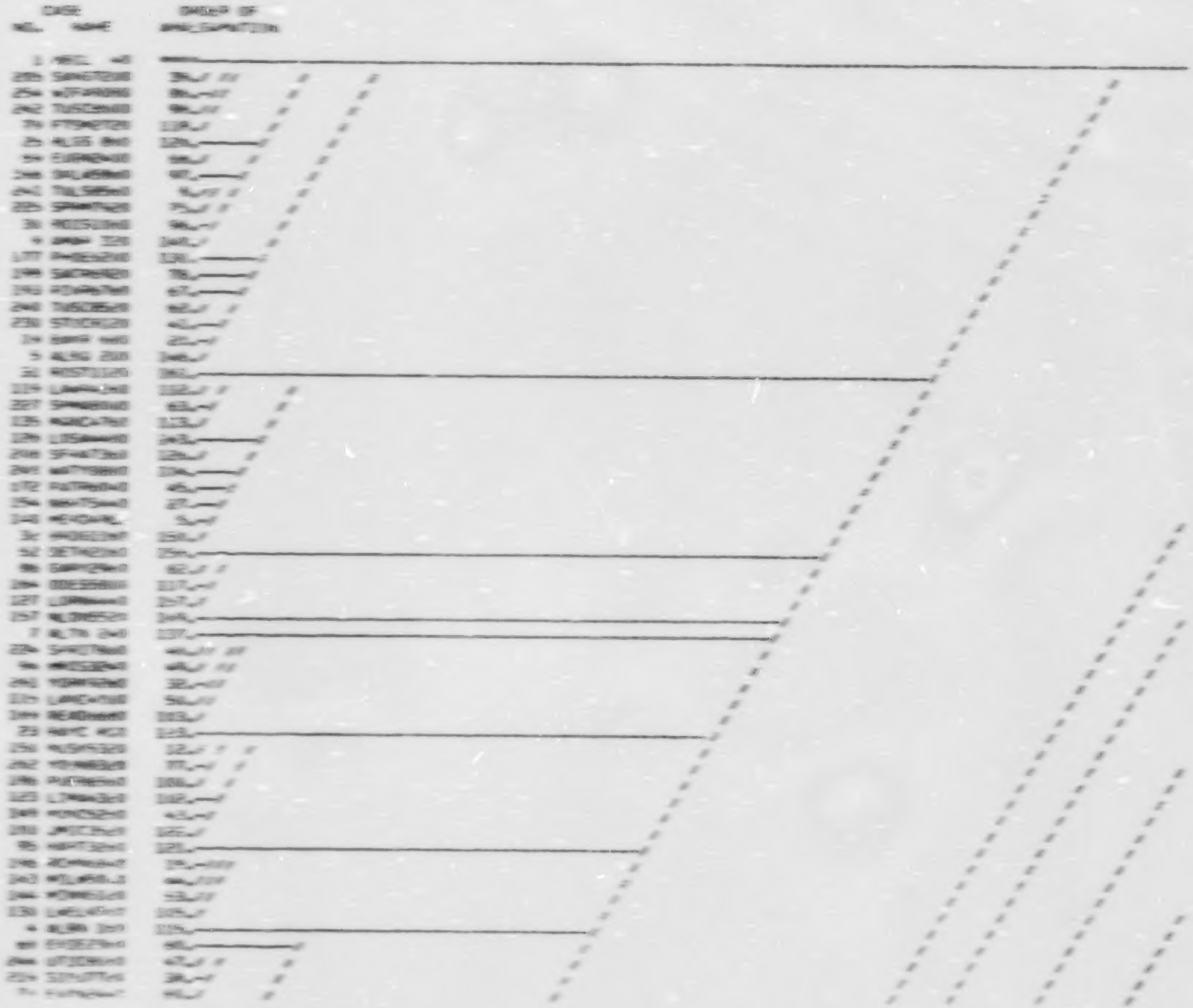
```
          EACH ROW CONTAINS THE VALUES OF VARIABLES FOR THE CLUSTER FORMED BY THE AMALGAMATION
           OF THE GIVEN ORDER. (THE LAST COLUMN REPRESENTS THE SUM OF CLUSTER CASE WEIGHTS)

AMALGAMATION
ORDER   DIST.
   1    0.512     -0.723      0.170     -0.633      0.067     -0.456      0.066      0.075      2.000
   2    0.476     -0.767      0.081     -0.641      0.073     -0.466     -0.036      0.017      3.000
   3    0.536      0.137      0.406     -1.040      0.543     -0.071      0.369     -1.386      2.000
   4    0.602     -1.765     -1.296     -0.266     -0.063      0.355     -0.792      0.362      2.000
   5    0.626     -0.908      0.510     -0.544      0.017      1.427     -0.193     -0.742      2.000
   6    0.641     -0.521     -0.126     -1.107      0.297     -0.121      0.486     -0.732      2.000
   7    0.644     -0.844      0.111     -0.511     -0.362      0.116      0.107      0.501      2.000
   8    0.686     -0.525     -0.096     -1.114      0.567     -0.134      0.460     -0.784      3.000
   9    0.702      0.465      0.498      0.176      0.465     -0.089     -0.423     -0.127      2.000
  10    0.744     -1.061      0.686     -1.250      0.696     -0.231      0.558      0.262      2.000
  11    0.773      0.130      0.544     -1.155      1.544     -0.151      0.466     -0.063      4.000
  12    0.778     -1.266     -0.453     -0.475      1.366     -0.472      1.061      0.045      2.000
  13    0.780      0.864      0.083     -0.526      0.698     -0.683     -0.242     -0.150      2.000
  14    0.787     -0.815     -0.075     -0.738      0.683     -0.750     -0.266     -0.134      3.000
  15    0.785     -0.941      0.615      1.427      0.476     -1.068     -1.145      0.686      2.000
  16    0.796     -0.515     -0.252     -1.063      0.411      0.245      0.067      0.294      2.000
  17    0.798     -0.572      0.163     -0.486     -1.187      0.065      0.255      0.567      3.000
  18    0.800     -1.276     -0.492      0.756     -1.635      0.195     -1.786     -2.052      2.000
  19    0.810     -0.845      0.422     -1.161     -0.595      0.448      0.415     -0.251      2.000
  20    0.811      0.429     -1.419      0.525      1.170      0.296     -0.326     -0.326      2.000
  21    0.820     -0.519     -0.063     -1.045      0.375     -0.077      0.425     -0.789      4.000
  22    0.818     -0.448     -0.006     -0.944      0.418     -0.253      0.096     -0.605      5.000
  23    0.822     -1.227     -0.152     -0.864     -0.456      0.077     -0.309     -0.796      2.000
  24    0.825      0.126      0.676     -0.561      0.195      0.268     -0.029      0.600      2.000
  25    0.829      1.292      0.261      0.756      1.031      0.612      0.616     -0.029      2.000
  26    0.843     -0.898     -0.227      1.652      0.538     -0.416      0.096      1.896      2.000
  27    0.863     -0.926      0.606     -0.591     -0.127      1.338     -0.236     -0.932      3.000
  28    0.876     -0.723     -0.032     -0.573     -0.298     -0.495     -0.176     -0.368      2.000
  29    0.876     -0.536      0.367      0.495  [illegible]   -1.171     -1.075      0.758      3.000
  30    0.876     -0.467      0.273      0.463  [illegible]   -1.020     -1.025      0.758      4.000
  31    0.881     -0.975     -0.097     -0.711     -0.377     -0.259     -0.193     -0.724      4.000
  32    0.885     -0.740      1.043     -0.596      0.267     -1.095     -0.295     -0.731      2.000
  33    0.887      0.260      0.067     -0.592     -0.012      0.156      0.019      0.552      3.000
  34    0.898     -0.092     -0.076     -0.926      0.000     -0.191      0.023      0.139      5.000
  35    0.900     -1.596     -0.305      0.429      0.273     -1.187     -1.147      0.311      2.000
  36    0.903      0.971      1.087      2.300     -0.133     -0.720      0.253      0.793      2.000
  37    0.935     -0.925     -0.230     -0.106     -0.793      0.007     -0.671     -0.796      2.000
  38    0.950     -0.726      0.796      0.267      0.447     -0.810     -0.519     -0.422      2.000
  39    0.952      0.963     -0.341      1.175      0.495     -1.161     -0.961     -1.089      2.000
  40    0.961      0.976      1.067     -0.467     -0.433      0.511      0.329      1.363      2.000
  41    0.962      0.240     -1.309      0.346      1.126      0.176     -0.356     -0.518      3.000
  42    1.010     -0.431     -0.051     -0.052      0.611     -0.397     -0.086      0.029      8.000
  43    1.019     -0.321     -0.117     -1.219      1.104     -0.737     -0.100      0.613      2.000
  44    1.026     -0.573      0.067     -1.159     -0.576      0.656      0.306     -0.325      3.000
  45    1.031     -0.865      0.673     -0.570     -0.095      1.522     -0.251     -0.793      4.000
  46    1.032      0.476      0.524      1.377     -0.656     -0.272      0.641      1.606      2.000
  47    1.033     -0.913     -0.751     -0.446      0.139     -0.937     -0.368     -0.444      2.000
  48    1.037     -0.017      1.116     -0.452     -0.095     -0.015     -0.429     -0.495      2.000
  49    1.037     -0.867      0.066     -0.405     -0.065     -0.055     -0.556     -1.296      3.000
  50    1.041     -0.046      1.037     -0.524      0.816     -1.191     -1.439     -0.473      5.000
  51    1.044     -0.900     -0.745      0.217      0.311     -1.021     -1.965      0.296      2.000
  52    1.050     -0.955     -0.241     -0.514     -0.515     -0.111     -0.419     -0.742      6.000
  53    1.060     -0.465      0.734     -1.194     -0.562      0.719      0.406     -0.306      4.000
  54    1.066     -0.296      0.351      0.996      0.572     -0.906     -0.461      0.725      5.000
  55    1.068     -1.023     -0.297     -0.136     -0.516     -0.867     -0.405     -0.635      7.000
  56    1.076     -0.323     -0.139      1.407      0.442     -0.729     -0.079      1.193      3.000
```

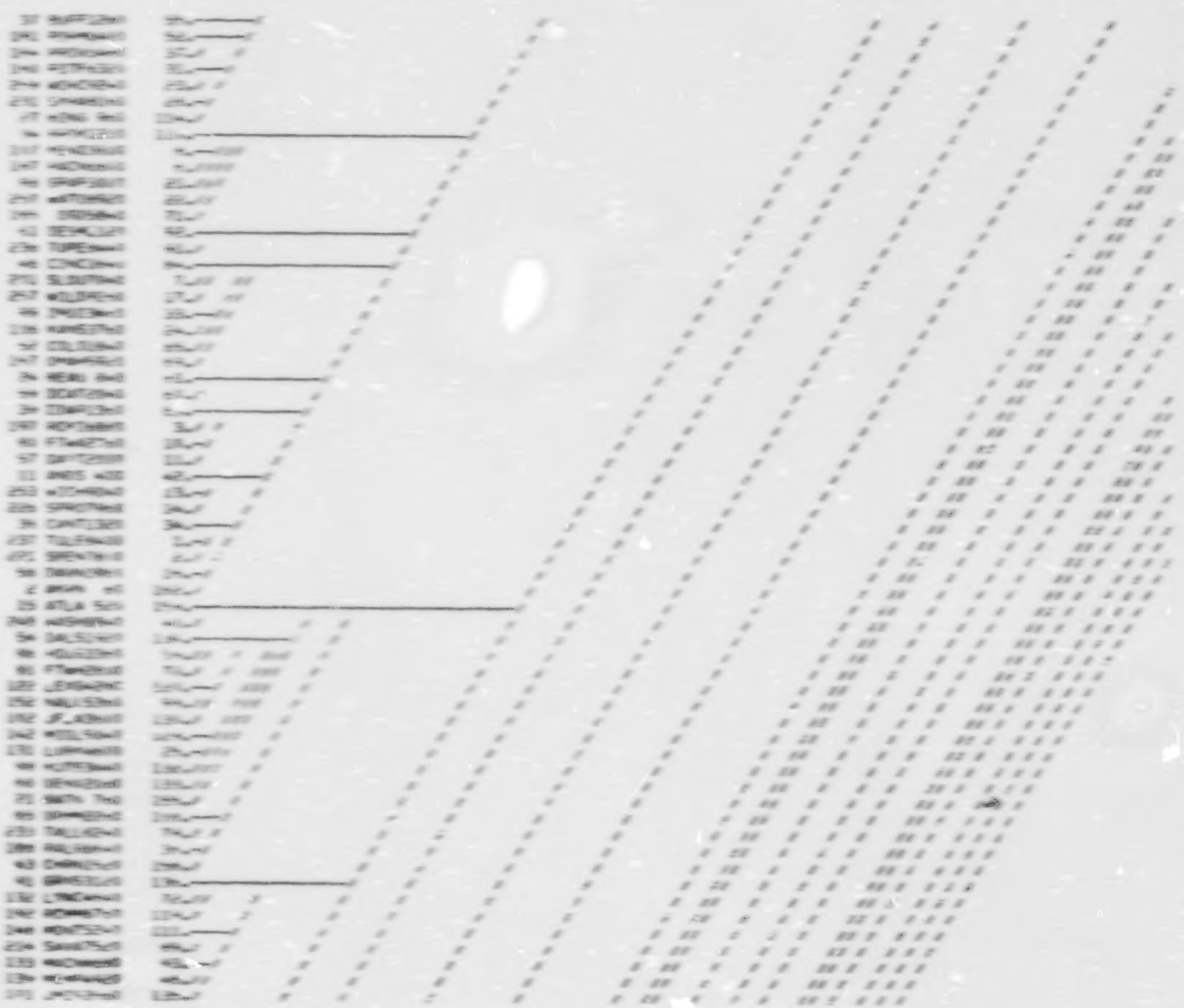
57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.016	1.944	2.056	2.000
58	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.441	-0.293	-1.577	2.000
59	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.237	0.323	0.876	2.000
60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.474	0.368	1.614	2.000
61	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.062	0.345	0.866	2.000
62	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.812	-0.475	-0.834	2.000
63	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.196	-0.508	-1.247	2.000
64	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.106	-1.546	0.300	4.000
65	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.110	0.137	0.576	6.000
66	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.325	0.097	-0.002	22.000
67	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.132	-0.406	-0.842	5.000
68	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.966	-0.433	-1.593	2.000
69	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.215	0.400	7.000
70	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.077	0.197	0.867	3.000
71	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.196	1.416	-0.907	6.000
72	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.135	-0.171	1.421	2.000
73	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.961	-0.367	0.001	2.000
74	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.107	-1.538	-1.899	3.000
75	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.295	-1.178	-0.121	3.000
76	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.360	1.760	-1.230	2.000
77	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.262	0.178	0.230	2.000
78	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.669	-0.525	-0.586	6.000
79	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.257	0.006	0.391	2.000
80	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.473	-0.443	-0.435	4.000
81	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.274	0.090	0.121	14.000
82	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.296	1.251	1.200	2.000
83	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2.543	0.176	-0.717	2.000
84	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.139	0.131	0.236	21.000
85	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.666	0.423	-1.425	2.000
86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.971	-0.779	0.101	2.000
87	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.097	-0.627	0.090	6.000
88	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.360	-0.416	-0.562	11.000
89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.011	-0.633	2.190	2.000
90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.442	-0.636	0.601	9.000
91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.363	-0.176	-0.518	2.000
92	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.368	0.246	0.172	23.000
93	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.309	0.694	1.713	3.000
94	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.003	0.683	1.774	2.000
95	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.020	1.091	1.036	3.000
96	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.804	-0.624	-1.127	2.000
97	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.417	-0.946	-0.522	5.000
98	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.286	-0.870	-0.646	4.000
99	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.356	-0.547	0.280	2.000
100	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.255	0.621	0.342	4.000
101	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.744	0.425	1.406	2.000
102	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.477	0.044	0.451	3.000
103	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.152	-0.625	-0.396	6.000
104	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.736	-0.381	-0.767	5.000
105	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.741	0.062	-0.477	5.000
106	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.485	0.956	1.767	4.000
107	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.148	0.371	1.201	2.000
108	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.341	0.079	0.542	4.000
109	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.710	0.087	1.237	4.000
110	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.268	0.243	-0.852	24.000
111	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.590	0.406	1.987	5.000
112	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.006	0.176	-1.380	2.000
113	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.596	-0.214	-1.323	4.000
114	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.376	-0.256	-0.535	12.000
115	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.229	0.102	-0.193	41.000
116	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.731	1.053	[illegible]	2.000
117	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.476	1.486	0.422	4.000

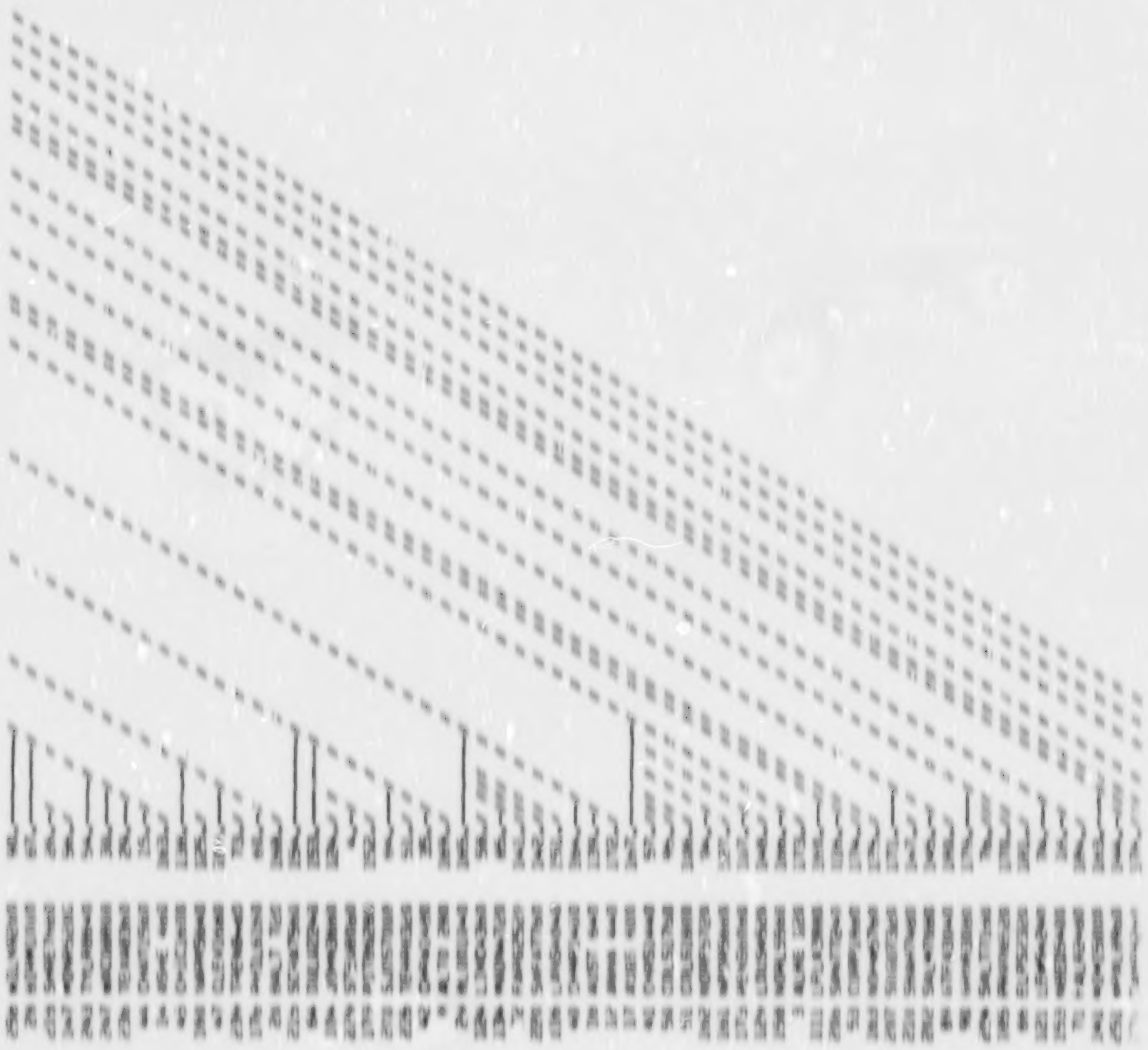
118	1.499	1.060	-0.021	1.225	0.025	-1.130	-0.987	-3.638	5.000
119	1.505	-0.362	1.471	0.540	-0.104	-0.736	-0.203	1.991	3.000
120	1.510	1.264	0.740	1.037	0.040	0.331	-0.436	0.530	3.000
121	1.514	-0.433	0.208	-0.717	0.257	-0.124	0.161	-0.224	40.000
122	1.540	-0.754	-0.706	-0.609	1.130	-0.507	0.375	0.275	7.000
123	1.550	-0.675	0.042	-0.730	0.293	-0.175	0.109	-0.150	53.000
124	1.560	-1.742	-1.587	-0.219	-0.467	-1.136	-0.843	-0.629	2.000
125	1.558	-0.766	0.676	-0.476	-1.018	2.025	0.411	1.132	2.000
126	1.580	0.346	1.294	-0.744	-0.084	1.629	-1.215	-0.261	2.000
127	1.585	1.794	-1.053	-0.654	-1.134	-0.644	0.438	1.706	6.000
128	1.592	0.762	0.336	0.206	0.481	-0.499	-0.482	-0.028	7.000
129	1.579	1.094	0.040	0.827	1.256	0.829	0.581	-0.222	3.000
130	1.595	1.325	0.444	0.567	0.308	0.704	-0.119	0.690	6.000
131	1.608	0.872	-1.091	0.467	1.064	0.081	-0.486	-0.541	7.000
132	1.544	1.213	0.877	0.567	0.060	0.479	0.190	0.392	9.000
133	1.551	1.747	0.657	0.449	0.661	0.514	0.477	0.299	10.000
134	1.629	1.835	0.402	0.447	0.632	0.440	0.177	0.372	11.000
135	1.665	-0.957	0.741	1.067	-0.288	-0.645	0.176	1.708	6.000
136	1.651	-0.265	0.454	1.144	0.096	-0.823	-0.253	1.329	17.000
137	1.673	-0.517	1.136	-0.717	0.254	-0.276	0.107	-0.192	54.000
138	1.674	-0.494	1.409	-0.407	-0.938	1.144	2.099	1.282	6.000
139	1.675	2.402	1.094	-0.840	-0.736	0.286	0.418	-0.336	2.000
140	1.678	0.717	-1.086	0.427	0.040	-0.739	-0.458	-0.685	12.000
141	1.694	1.796	1.385	-0.427	-0.447	-0.479	0.325	-1.561	4.000
142	1.683	0.876	1.324	-0.316	-0.404	-1.421	0.434	-1.575	5.000
143	1.705	-0.489	0.472	-0.404	-0.336	1.098	-0.562	-0.609	7.000
144	1.741	0.804	-0.388	1.043	1.308	1.920	-2.254	0.376	2.000
145	1.750	0.745	-2.267	-0.272	0.061	-0.538	-1.133	-1.058	2.000
146	1.771	0.068	-0.431	0.567	0.427	-0.245	-0.728	-0.638	15.000
147	1.779	1.747	-0.402	-0.568	-0.825	1.040	-0.670	-0.758	2.000
148	1.826	0.857	-1.125	-0.089	-0.746	-0.525	0.790	1.639	8.000
149	1.840	-0.640	0.110	-0.708	0.246	-0.277	0.106	-0.104	60.000
150	1.844	-0.425	0.357	-0.365	-0.767	1.294	-0.435	-0.080	11.000
151	1.840	0.467	1.491	-0.444	-0.655	-0.927	0.541	-1.426	6.000
152	1.842	-1.733	-1.336	-0.257	-0.275	-0.390	-0.716	-0.134	4.000
153	1.703	-0.442	-0.444	0.131	0.600	-0.767	-1.132	0.083	6.000
154	1.887	-1.536	-0.494	0.124	-0.043	-0.690	-1.232	0.083	4.000
155	1.960	1.413	0.947	0.756	0.444	0.588	0.408	0.521	13.000
156	2.024	1.947	1.136	0.645	1.278	0.164	0.172	0.538	17.000
157	2.036	-0.472	-0.032	-0.725	1.146	0.032	1.073	0.839	4.000
158	2.029	-0.644	0.337	-0.714	0.385	-0.208	0.167	-0.129	64.000
159	2.037	0.441	0.747	0.740	0.260	-0.337	-0.841	0.933	34.000
160	2.104	0.313	-0.285	1.542	0.645	0.108	1.408	-0.138	3.000
161	2.144	-0.504	0.346	-0.449	0.174	0.013	0.178	-0.236	75.000
162	2.134	-0.273	0.024	-0.427	0.265	-0.034	-0.083	-0.313	94.000
163	2.157	-0.204	0.254	-0.241	0.254	-0.114	-0.072	0.016	126.000
164	2.242	2.404	0.702	-0.704	-1.067	-0.184	0.276	-0.359	3.000
165	2.133	1.341	1.095	-0.544	-0.742	-0.602	0.453	-1.082	4.000
166	2.234	0.274	-0.244	1.465	-1.567	-0.181	0.450	1.074	2.000
167	2.113	0.724	-0.494	0.259	-0.493	-0.452	0.482	1.527	10.000
168	2.262	-0.110	0.044	-0.125	0.208	-0.061	-0.044	0.071	134.000
169	2.241	-0.210	0.172	-0.108	0.243	-0.181	-0.141	0.067	143.000
170	2.344	0.274	-0.008	-0.569	0.367	0.031	2.131	-1.577	3.000
171	2.346	1.245	-0.829	-0.421	0.021	0.251	-0.402	-1.304	4.000
172	2.545	-0.174	0.225	-0.124	0.125	-0.135	-0.106	-0.001	152.000
173	2.573	-0.046	0.252	-0.105	0.054	-0.155	-0.040	0.043	162.000
174	2.629	-0.548	0.347	-0.104	0.066	-0.149	-0.117	0.089	165.000
175	2.713	-0.544	0.256	-0.076	0.071	-0.137	-0.017	0.099	166.000
176	2.719	0.476	-0.838	2.053	0.405	2.443	-2.178	0.259	3.000
177	2.624	-1.936	-0.294	0.967	0.849	1.082	3.433	-1.419	2.000
178	2.444	-1.797	-0.817	0.432	[illegible]	2.717	-1.103	0.445	2.000

179	2.435	0.124	-1.136	-0.046	0.132	0.396	2.109	-1.492	4.000
180	2.494	-0.217	0.109	-0.086	0.070	-0.127	-0.037	0.057	170.000
181	3.019	-0.714	0.000	-0.104	0.071	-0.115	0.013	0.021	174.000
182	3.146	2.746	-1.419	-1.110	-0.420	-0.076	0.513	1.111	3.000
183	3.199	0.068	-0.173	7.852	0.525	2.013	-2.161	0.490	4.000
184	3.447	-0.832	0.070	-0.071	0.034	-0.112	-0.013	-0.012	177.000
185	3.657	-0.731	0.085	-0.029	0.040	-0.035	-0.060	-0.081	181.000
186	3.829	-0.735	0.071	-0.037	0.056	-0.064	-0.041	0.000	182.000
187	3.836	-0.714	0.073	-0.040	0.069	-0.040	-0.039	-0.005	184.000
188	5.025	0.051	0.014	-0.054	0.053	-0.041	-0.030	0.014	187.000
189	6.111	0.026	0.014	-0.052	-0.001	-0.011	-0.041	0.029	189.000
190	6.933	0.005	-0.000	0.000	-0.004	0.000	0.000	0.000	191.000

This table narrates the grouping process. At the beginning, each observation is considered a group. The algorithm locates the two groups with centroids closest to each other and combines them into one group. This is indicated as step 1 in the column at the left. The second column indicates the centroids of these two groups were located at a distance of .512 from each other. The third through ninth columns locate the centroid of the newly formed group along each of the seven axes and the last column indicates the number of observations in the new group (given uniform weights for all cases). The second row describes the second step which groups together the two groups which are closest following the formation of the first group. The process is repeated until all cities have been combined into one group. The results are interpreted with the aid of the tree diagram below.


CASE
NO. NAME
ORDER OF
AMALGAMATION





CONTINUATION OF [illegible] TAPE

This diagram is read like a branching diagram. Beginning at the upper right, follow the horizontal line to the left until a diagonal branches off. At this point, the diagonal divides the SMSA list at the left into two subgroups. These subgroups are further divided by other horizontal and diagonal lines all the way down to groups of one and two. To see which groups had formed at a particular point in the grouping process, consult the column labeled "order of amalgamation." In this diagram, for instance, there are 191 cities. To see what cities had grouped together by step 71, consider all lines emanating from numbers greater than 70 isolated and look at the remaining tree.